PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
A61K 39/12, C07K 3/00
C07H 15/12

A1

(11) International Publication Number: WO 93/15763

(43) International Publication Date: 19 August 1993 (19.08.93)

(21) International Application Number: PCT/US93/01451

(22) International Filing Date: 18 February 1993 (18.02.93)

(30) Priority data:
07/837,773 18 February 1992 (18.02.92) US

(60) Parent Application or Grant
(63) Related by Continuation
US 07/837,773 (CON)
Filed on 18 February 1992 (18.02.92)

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, P.O. Box 7929, Philadelphia, PA 19101 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : SHATZMAN, Allan [US/US]; 332 Rees Drive, King of Prussia, PA 19406 (US). SCOTT, Miller [US/US]; 13 Eagle Stream Drive, Apt. Y-18, Norristown, PA 19402 (US). DILLON, Susan, B. [US/US]; 17 Raven Drive, Chadds Ford, PA 19317 (US).

(74) Agent: SUTTON, Jeffrey, A.; SmithKline Beecham Corporation, Corporate Patents - U.S., UW2220, 709 Swedeland Road, P.O. Box 1538, King of Prussia, PA 19406-0939 (US).

(81) Designated States: AU, CA, JP, KR, NZ, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: VACCINAL POLYPEPTIDES

(57) Abstract

This invention provides vaccine compositions capable of conferring multi-strain immunity against influenza A and influenza B.

18912
#703

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NL**	Netherlands
BE	Belgium	**GN**	Guinea	**NO**	Norway
BF	Burkina Faso	**GR**	Greece	**NZ**	New Zealand
BG	Bulgaria	**HU**	Hungary	**PL**	Poland
BJ	Benin	**IE**	Ireland	**PT**	Portugal
BR	Brazil	**IT**	Italy	**RO**	Romania
CA	Canada	**JP**	Japan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SK**	Slovak Republic
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SN**	Senegal
CM	Cameroon	**LI**	Liechtenstein	**SU**	Soviet Union
CS	Czechoslovakia	**LK**	Sri Lanka	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**MC**	Monaco	**UA**	Ukraine
DK	Denmark	**MG**	Madagascar	**US**	United States of America
ES	Spain	**ML**	Mali	**VN**	Viet Nam
FI	Finland	**MN**	Mongolia		

VACCINAL POLYPEPTIDES

This is a continuation-in-part of pending United States patent application Serial Number 751,896; which is a continuation-in-part of United States patent application Serial Number 387,558; which is a continuation-in-part of United States patent application Serial Number 238,801, now abandoned; which is a continuation-in-part of United States patent application Serial Number 645,732, now abandoned.

Field of the Invention

The present invention relates generally to a polypeptide useful in a composition for providing immunity against influenza A and influenza B in an animal.

Background of the Invention

Influenza virus infection causes acute respiratory disease in man, horses, swine and fowl, sometimes of pandemic proportions. Influenza viruses are orthomyxoviruses and, as such, have envelope virions of 80 to 120 nanometers in diameter, with two different

glycoprotein spikes. Three types, A, B and C, infect humans. Type A viruses have been responsible for the majority of human epidemics in modern history, although there are also sporadic outbreaks of Type B infections. Known swine, equine and avian viruses have mostly been Type A, although Type C viruses have also been isolated from swine.

The Type A viruses are divided into subtypes based on the antigenic properties of the hemagglutinin (HA) and neuraminidase (NA) surface glycoproteins. Within type A, subtypes H1 ("swine flu"), H2 ("asian flu") and H3 ("Hong Kong flu") are predominant in human infections. In swine, the predominant influenza A subtypes are H1 and H3; in horses, H3 and H7; and in avians, H5 and H7. Presently only one Type B virus has been identified, with no subtypes.

Genetic "drift" or "shift", i.e., rapid and unpredictable change in the antigen, occurs at approximately yearly intervals, and affects antigenic determinants in the HA and NA proteins. Therefore, it has not been possible to prepare a "universal" influenza virus vaccine using conventional killed or attenuated viruses, that is, a vaccine which is non-strain specific.

Recently, attempts have been made to prepare such universal, or semi-universal, vaccines from reassortant viruses prepared by crossing different strains. More recently, such attempts have involved recombinant DNA techniques focusing primarily on the HA protein.

There remains a need in the art for vaccine formulations and compositions capable of inducing protective responses in animals against influenza viruses.

Summary of the Invention

The present invention provides compositions containing, and methods for use of, a protein which is capable of inducing protection in animals and avians against challenge with more than one strain of influenza type A and influenza type B.

Thus, one aspect of the invention provides a DNA sequence encoding a modified purified recombinant protein. The DNA sequence of the invention encodes a modified protein sequence derived from the HA2 subunit of a selected hemagglutinin (HA) protein. In one embodiment, the sequence is derived from an H3N2 subtype influenza virus. These H3N2 fusion proteins are capable of inducing T cell responses in the absence of

neutralizing antibodies. In another embodiment, a DNA sequence of this invention encodes a modified protein sequence derived from the HA2 subunit from a type B influenza virus. Still further embodiments include DNA sequences obtained as described for the two above virus, where the sequences are derived from other Type A influenza strains infecting animals as well as humans. Such virus include, without limitation, Type A subtypes of H1, H2, H3, H4, H5, H6 and H7.

In another aspect, the invention provides a DNA sequence encoding a recombinant fusion protein, in which the desired Type A subtype HA2 subunit sequence or a portion thereof, is fused in frame to another protein or protein fragment capable of enhancing expression of the fusion protein. One embodiment includes the H3N2 subtype HA2 subunit sequence described above fused in frame to another protein or fragment capable of enhancing expression thereof. Another embodiment of such a fusion protein comprises a type B HA2 sequence, described above, or a portion thereof, fused in frame to another protein or protein fragment capable of enhancing expression of the fusion protein. Still other Type A subtype HA2 sequences can be similarly used. It is desirable that this fusion partner protein be an influenza protein sequence or fragment thereof.

In still another aspect a protein encoded by a DNA sequence of the invention is provided. The protein may be a protein sequence derived from the HA2 subunit of a hemagglutinin (HA) protein from a selected Type A subtype virus. Desirably the subtype virus is an H3N2. In another embodiment, the protein may be derived from the HA subunit from a type B influenza virus. Other embodiments include H5 or H7 subtypes. Additionally, preferred embodiments include fusion proteins comprising a protein sequence derived from the HA2 subunit of an HA protein from a Type A virus, e.g., an H3N2 subtype, or from a type B virus fused in frame to a selected influenza sequence. The proteins of this invention are particularly useful in inducing protection in mammals, especially humans, against challenge by type B or an H3N2 subtype of influenza A. The proteins employing other Type A subtypes, e.g., H5 and H7, are useful in inducing protection in animals against influenza viruses.

In a further aspect the invention provides a vaccine composition containing a purified protein of the invention, as described above. Such a vaccine composition may include a fusion protein of the invention. In other embodiments of the invention, the vaccine compositions contain an H3HA2 protein of the invention and other influenza antigens; a type B HA2

protein of the invention and other influenza antigens; or both an H3HA2 protein, a BHA2 protein and other influenza antigens. In a preferred embodiment for human use, a combination vaccine of the invention will contain an H3HA2 and a BHA2 protein of the invention in combination with influenza antigens derived from the other type A influenza virus subtypes, H1 and H2. An embodiment for use in animals may contain an H5HA2 or H7HA2 protein, among others.

A further aspect of this invention is a method for inducing in an animal protection against influenza type A, influenza type B, influenza type C, or combinations thereof, which comprises internally administering to the animal an effective immunogenic amount of a vaccine composition of the present invention.

Still a further aspect of this invention is a method for inducing in an animal protection against multiple strains of influenza types A and B which comprises internally administering to the animal an effective immunogenic amount of a vaccine composition of the present invention.

Other aspects and advantages of the present invention are described further in the following detailed description of the preferred embodiments thereof.

Brief Description of the Drawings

Fig. 1 illustrates the nucleic acid sequences of the HA2 portions of (a) A/Udorn [SEQ ID NO: 1], (b) A/Victoria [SEQ ID NO: 3], (c) A/PR/8/34 [SEQ ID NO: 5], and (d) a consensus sequence [SEQ ID NO: 7]. Dashes indicate the same nucleotide as the consensus sequence. Different nucleotides from that of the consensus sequence are reported in lower case letters. Dots indicate no corresponding nucleotide when compared to the consensus sequence.

Fig. 2 illustrates the nucleic acid and amino acid sequences of $NS1_{(1-81)}H3HA2_{(1-221)}$ fusion protein [SEQ ID NO: 9 & 10].

Fig. 3 illustrates the nucleic acid and amino acid sequences of the $NS1_{(1-81)}H3HA2_{(77-221)}$ fusion protein [SEQ ID NO: 11 & 12].

Fig. 4 illustrates the nucleic acid and amino acid sequences of the type B fusion protein, $NS1_{1-42}HA2_{41-223}$. [SEQ ID NO: 13 & 14].

Detailed Description of the Invention

The present invention provides novel proteins, DNA sequences, pharmaceutical vaccine compositions and methods of use thereof for conferring protection in vaccinated mammals against one strain, or desirably

multiple strains, of influenza viruses. The proteins and vaccine compositions of the present invention demonstrate the ability to stimulate or produce a protective immune response which is capable of recognizing an influenza virus or influenza virus-infected cells and protecting the vaccinated mammal against disease caused thereby. This protective response is desirably a T cell response, produced in the substantial absence of vaccine-induced neutralizing antibody.

While the proteins and DNA sequences specifically described herein are directed to the H3HA2 and BHA2 sequences originating from viral strains to which humans are susceptible, it is expected that similar sequences and molecules can be prepared for veterinary applications. For example, selected HA2 sequences obtained from type A viral strains, e.g., H5HA2, H7HA2 and other strains of interest may be obtained following the teachings described herein for the exemplified H3HA2 and BHA2 sequences. One of skill in the art should understand that this invention is not limited to the exemplified protein and DNA sequences, even though the following disclosure is limited to the two latter sequences for simplicity. Such additional viral HA2 subunits are expected to share the biological characteristics of the exemplified sequences.

Thus, this invention provides a protein or fragment thereof characterized by an amino acid sequence derived from the HA2 subunit of a hemagglutinin (HA) protein, e.g., from a H3N2 subtype virus. The H3 proteins of the invention are capable of inducing T helper cells, particularly cytotoxic T lymphocytes, in the absence of neutralizing antibodies. Among H3N2 subtype strains of influenza A include A/Udorn and A/Victoria viruses. Other H3N2 virus strains of influenza A may also produce HA proteins for use in vaccine compositions according to this invention. Fig. 1 compares the nucleic acid sequences of the HA2 portions of the A/Udorn [SEQ ID NO: 1] and A/Victoria [SEQ ID NO: 3] strains with the nucleic acid sequence of an H1N1 subtype virus, A/PR/8/34 [SEQ ID NO: 5]. A consensus sequence [SEQ ID NO: 7] was computer generated, and may likewise be useful in producing proteins according to this invention. This consensus sequence [SEQ ID NO: 7] can be constructed by a commercially available computerized sequence analysis program, such as Genetics Computers Group [Univeristy of Wisconsin].

Proteins according to this invention may include unfused HA2 subunits of the influenza A viruses, particularly H3N2 subtype. For example, in one embodiment, a protein of the invention contains amino

acids 1-221 of a selected H3HA2 subunit. In another embodiment, a protein of the invention contains amino acids 77-221 of the H3HA2 subunit. Other fragments of this HA2 amino acid sequence characterized by the ability to stimulate similar immunological activity in an immunized animal are also encompassed by this invention.

Proteins of this invention also include fusion proteins comprising a protein sequence derived from the HA2 subunit of an HA protein from a Type A virus, e.g., an H3N2 subtype virus, fused in frame to another protein or protein fragment capable of enhancing expression of the fusion protein. It is desirable that this fusion "partner" protein be an influenza protein sequence or fragment thereof derived from the same or another strain of influenza virus as the HA protein or protein fragment. Preferably, this fusion partner protein is all or a portion of the influenza virus NS1 gene or an HA2 subunit.

In the embodiments exemplified herein, the NS1 portion of the fusion protein is derived from an H1N1 subtype virus, A/PR/8/34. For example, in one embodiment, the NS1 portion may comprise amino acid residues 1 to 42 of H1NS1. In another embodiment the NS1 portion may comprise amino acid residues 1 to 81 of the selected virus. The HA2 fragment may alternatively be fused to a portion of the NS1 peptide derived from a

selected Type A virus, e.g., an H3 subtype virus (H3HA2), or a type B (BHA2) virus.

However, other non-influenza fusion proteins may also produce desirable fusion proteins with the H3N2, or other Type A, or type B protein or portion thereof. Thus, in still another alternative embodiment, as discussed below, the HA2 fragment may be fused to any peptide capable of enhancing its expression in the host cell selected. One of skill in the art may readily select a fusion "partner" protein or fragment taking into account the desired host cell and utilizing the teachings herein. The fusion proteins of the present invention are not limited by the selection of the "partner" protein or fragment to which the HA2 fragment is fused.

In yet another embodiment, the present invention provides a modified protein containing a portion of the HA2 subunit of a type B influenza virus. Currently, the preferred human virus strain is B/Lee/40. However, the vaccinal proteins of this invention are not limited to this type B strain, and other strains infecting other species, or other as yet unidentified type B virus strains, may be used to produce the HA2 protein. These type B HA2 proteins may be fused, as described above for the H3HA2 proteins of this invention, or remain unfused.

In the construction of a fusion protein according to this invention, a linker sequence may be inserted optionally between the two fused sequences, i.e., between the NS1 portion and the HA2 portion. This optional linker may provide space between the two linked sequences. Alternatively, this linker sequence may encode, if desired, a polypeptide which is selectively cleavable or digestible by conventional chemical or enzymatic methods. For example, the selected cleavage site may be an enzymatic cleavage site, including sites for cleavage by a proteolytic enzyme, such as enterokinase, factor Xa, trypsin, collagenase and thrombin. Alternatively, the cleavage site in the linker may be a site capable of being cleaved upon exposure to a selected chemical, e.g., cyanogen bromide or hydroxylamine. The cleavage site, if inserted into a linker useful in the fusion sequences of this invention, does not limit this invention. Any desired cleavage site, of which many are known in the art, may be used for this purpose.

A presently preferred example of a fusion protein of this invention is $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10], which comprises the first 81 amino acids of NS1 fused to amino acid 1 to 221 of the H3HA2 subunit (amino acids 1-221). Another exemplary fusion protein, $NS1_{(1-81)}H3HA2_{(77-221)}$ [SEQ ID NO: 12], comprises the first 81 amino

acids of NS1 fused to amino acid 77 to 221 of the truncated H3HA2 subunit. Yet another preferred example of a fusion protein of this invention is $NS1_{1-42}BHA2_{41-223}$ [SEQ ID NO: 14], which comprises the first 42 amino acids of NS1 fused to amino acids 41 to 223 of the truncated BHA2 subunit. These proteins, fusion proteins and similar proteins encoded by the below-described DNA sequences are referred to collectively herein as H3HA2 proteins.

The $NS1_{(1-81)}H3HA2_{(1-221)}$ protein [SEQ ID NO: 10] of the invention has a three-dimensional structure which is substantially similar to that of the $NS1_{(1-81)}HA2_{(1-222)}$ protein [SEQ ID NO: 16] derived from the H1N1 subtype virus (C13). However, the amino acid sequence of the $NS1_{(1-81)}H3HA2_{(1-221)}$ protein [SEQ ID NO: 10] has only approximately 50% homology with the amino acid sequence of C13 protein [SEQ ID NO: 16]. Additionally, as illustrated in Fig. 1, the nucleic acid sequence of the $H3HA2_{1-221}$ fragment derived from A/Udorn (nucleotides 25-560 from that virus) [SEQ ID NO: 1] has only approximately 60% homology with the nucleic acid sequence of the $H1HA2_{1-222}$ protein derived from strain A/PR/8/34 (nucleotides 1872-2407 from A/PR/8/34) [SEQ ID NO: 5]. However, the nucleic acid sequence of $H3HA2_{1-221}$ from A/Udorn (nucleotides 1-499 of A/Udorn) [SEQ ID NO: 1] has approximately 99% homology with the nucleic acid sequence of $H3HA2_{1-221}$ from A/Victoria/H3/75

(nucleotides 1226-1725 of A/Victoria) [SEQ ID NO: 3] [Fiers et al, Cell, 19:683-696 (1980)].

Analogs of the HA2 peptides from a Type A virus, e.g., an H3, or B viruses, included within the definition of this invention, include truncated polypeptides (including fragments) and HA2 polypeptides, e.g. mutants that retain the epitopes and thus the biological activity of HA2. It is anticipated that, because the NS1 portion of the fusion peptide provides a means of expressing the protein at high levels and does not appear to play as significant a role in the immunological responses to the HA2 fusion proteins as does the HA2 portion, any number of analogs of this fusion partner can be made.

Typically, the analogs of the HA2 peptides and/or the fusion partner differ by only 1 to about 4 codon changes. Other examples of analogs include polypeptides with minor amino acid variations from the natural amino acid sequence of HA2; in particular, conservative amino acid replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are generally divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) non-polar = alanine, valine, leucine, isoleucine, proline,

phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid will not have a significant effect on its activity, especially if the replacement does not involve an amino acid at an epitope of the HA2 polypeptide. The construction of such analogs, given the description herein and conventional methods of protein modification known to one of skill in the art, are believed to be encompassed by this invention.

Currently, it is theorized that the HA2 portion of the fusion peptide (e.g., $H3HA2_{1-221}$, $H3HA2_{77-221}$ and $BHA2_{41-223}$) confers the majority of the necessary epitopes for antibody binding or T cell (particularly CTL) targeting. Once these epitope sequences are precisely identified, portions of the HA2 sequence which are not part of these epitopes may be altered without significantly affecting the bioactivity of the fusion protein.

The present invention also encompasses DNA sequences of this invention encoding the above-described proteins and fusion proteins, the sequences characterized by having an immunogenic determinant of a modified HA2 subunit of an HA protein, derived from a Type A virus, e.g., an H3 subtype, or type B virus. Other DNA sequences of this invention encode such HA2 subunits, optionally fused to a DNA sequence encoding a protein or peptide which is capable of enhancing expression of the protein in a selected host cell. For example, the consensus sequence illustrated in Fig. 1(d) may provide a source of HA2 DNA. The currently preferred embodiment provides a DNA sequence encoding a Type A virus, e.g., an H3 or type B HA2 protein or fragment thereof fused in frame to a DNA sequence encoding a portion of the nonstructural influenza protein 1 (NS1).

Coding sequences for the HA2, NS1 and other viral proteins of influenza virus can be prepared synthetically or can be derived from viral RNA or from available cDNA-containing plasmids by known techniques. For example, in addition to the above-cited references, a DNA coding sequence for HA from the A/Japan/305/57 strain was cloned, sequenced and reported by Gething et al, Nature, 287:301-306 (1980). An HA coding sequence for strain A/NT/60/68 was cloned as reported by Sleigh et al, and by Both et al, in Developments in Cell Biology,

Elsevier Science Publishing Co., pages 69-79 and 81-89, respectively, (1980). An HA coding sequence for strain A/WSN/33 was cloned as reported by Davis et al, Gene, 10:205-218 (1980); and by Hiti et al, Virology, 111:113-124 (1981). An HA coding sequence for fowl plague virus was cloned as reported by Porter et al and by Emtage et al, both in Developments in Cell Biology, cited above, at pages 39-49 and 157-168. Also, influenza viruses, including other strains, subtypes and types, are available from clinical specimens and from public depositories, such as the American Type Culture Collection (ATCC), Rockville, Maryland, U.S.A.

Allelic variations (naturally-occurring base changes in the species population which may or may not result in an amino acid change) of DNA sequences encoding the H3HA2 or BHA2 protein sequences are also included in the present invention, as well as analogs or derivatives thereof. Similarly, DNA sequences which code for H3 or other Type A or type B HA2 proteins of the invention but which differ in codon sequence due to the degeneracies of the genetic code or variations in the DNA sequence encoding H3HA2, other Type A or BHA2 proteins which are caused by point mutations or by induced modifications to enhance the activity, half-life or production of the peptide encoded thereby are also encompassed in the invention. Also covered by this invention are DNA

sequences which hybridize under stringent conditions with the DNA sequences encoding the HA2 subunit proteins, e.g., H3HA2 or BHA2 proteins, of this invention. DNA sequences which hybridize under non-stringent conditions with the disclosed sequences, but which encode proteins or fragments retaining the biological activities of the H3HA2 or BHA2 proteins, are also included in this invention. Typical conditions for stringent or non-stringent hybridization are known to those of skill in the art. [See, e.g., Sambrook et al, Molecular Cloning. A Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory, NY (1989)].

The fusion proteins of the invention may be prepared by conventional genetic engineering and recombinant techniques known to those of skill in the art. Similarly, the proteins may be purified from expression in host cell or vector systems by conventional means.

Systems for cloning and expression of the vaccinal polypeptide of this invention in various microorganisms and cells, including, for example, E. coli, Bacillus, Streptomyces, Saccharomyces, mammalian and insect cells, are known and available from private and public laboratories and depositories and from commercial vendors. The preferred host is E. coli because it can be used to produce large amounts of

desired proteins safely and cheaply. The polypeptide employed in the presently preferred embodiment is expressed in E. coli. To circumvent the requirement of ampicillin for plasmid selection in production fermentations, a preferred method of production employs an alternative expression system in which the β-lactamase coding sequence is wholly or partially replaced by a coding sequence for an alternative selectable marker such as, for example, kanamycin or chloramphenicol.

To aid in expression of the H3 or other Type A subunit or type B HA2 peptides or fusion protein described above, these protein sequences or fragments thereof may also be fused to a polypeptide capable of enhancing expression of these fragments in the selected host system. Ordinarily, such a peptide would contain a leader sequence fragment that provides for secretion of the Type A subunit fragment, e.g., the H3HA2 fragment, or type B HA2 fragment in the host cell. The leader sequence fragment typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. There may be processing sites encoded between the leader sequence and the Type A subtype or type B HA2 fragment that can be cleaved either in vivo or in vitro. Alternatively, a promoter sequence may be linked directly with the DNA molecule encoding the HA2 fragment. Such polypeptides,

promoter and leader sequences are known to those of skill in the art and may be readily selected for expression in the selected host.

Construction of expression systems, including expression vectors and transformed host cells are thus within the art. See, generally, methods described in standard texts, such as Sambrook et al, Molecular Cloning A Laboratory Manual, 2d edit., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989). The present invention is therefore not limited to any particular expression system or vector, nor to any particular purification process from cell lysates or cell medium.

The proteins and fusion proteins of this invention may be employed in vaccine compositions. Pharmaceutical vaccine compositions of this invention, therefore, contain an effective immunogenic amount of a selected HA2 protein, e.g., H3HA2 or BHA2 protein, of the invention in admixture with a suitable adjuvant in a nontoxic and sterile pharmaceutically acceptable carrier.

Suitable carriers for vaccine use are well known to those of skill in the art. However, exemplary carriers include sterile saline, lactose, sucrose, calcium phosphate, gelatin, dextrin, agar, pectin, peanut oil, olive oil, sesame oil, squalene and water. Additionally, the carrier or diluent may include a time delay material, such as glyceryl monostearate or glyceryl

distearate alone or with a wax. Optionally, suitable chemical stabilizers may be used to improve the stability of the pharmaceutical preparation. Suitable chemical stabilizers are well known to those of skill in the art and include, for example, citric acid and other agents to adjust pH, chelating or sequestering agents, and antioxidants.

While any aluminum adjuvant may be used in the vaccine compositions of this invention, two desirable adjuvants are commercially marketed under the trademarks Rehsorptar [Armour Pharmaceuticals, Kankakee, IL] and Rehydragel [Reheis Chemical Co., Berkeley Heights, NJ]. These products are aluminum hydroxide gels which contain approximately 2% w/v Al_2O_3, which is equivalent to approximately 10.6 mg/r Al^{+3}.

Vaccine compositions of this invention may employ an immunogenic amount of a purified recombinant protein as described above. A preferred embodiment of the vaccine of the invention is composed of an aqueous suspension or solution containing the recombinant HA2 protein molecule, e.g., H3HA2 or BHA2, together with an adjuvant, preferably an aluminum, most preferably aluminum hydroxide, buffered at physiological pH, in a form ready for injection. A preferred protein for use in these vaccine compositions includes a protein comprising amino acid residues 1 to 81 from NS1 fused to C-terminal

amino acid residues 1-221 from the hemagglutinin subunit 2 (HA2) from influenza A, subtype H3N2. Another preferred vaccine composition of this invention employs a purified recombinant protein made up of amino acid residues 1 to 81 from NS1 fused to amino acid residues 77-221 of the HA2 from influenza A, subtype H3N2. Still another preferred vaccine composition of this invention employs a purified recombinant protein made up of amino acid residues 1 to 42 fused to amino acid residues 41-223 of the HA2 from influenza B.

Vaccine compositions of the invention may also employ an immunogenic amount of a recombinant protein of the invention in combination with other influenza antigens. Suitable influenza antigens for combination in a vaccine composition with the proteins of this invention may be derived from type A, H1 subtype viruses and may include the recombinant fusion proteins described in detail in copending U. S. Patent Application Ser. No. 07/387,200, filed July 28, 1989 and its corresponding European Patent Application No. 366, 238, published May 2, 1990; and in co-pending U. S. Patent Application Ser. No. 07/387,558, filed July 28, 1989 and its corresponding European Patent Application No. 366,239, published May 2, 1990. The C13 protein ($NS1_{(1-81)}HA2_{(1-222)}$) [SEQ ID NO: 15 & 16], D protein ($NS1_{(1-81)}HA2_{(65-222)}$) [SEQ ID NO: 17 & 18] and other fusion proteins derived from the H1N1 influenza

virus subtype and the recombinant expression and purification thereof are disclosed in detail in these applications, and in the parent applications identified in this application, all of which are incorporated by reference herein.

More specifically, suitable H1 subtype immunogenic proteins include C13 ($NS1_{(1-81)}$-D-L-S-R-$HA2_{(1-222)}$) [SEQ ID NO: 15 & 16], D ($NS1_{(1-81)}$-Q-I-P-$HA2_{(65-222)}$) [SEQ ID NO: 17 & 18], C13 short ($NS1_{(1-42)}$-M-D-L-S-R-$HA2_{(1-222)}$) [SEQ ID NO: 19 & 20], D short ($NS1_{(1-42)}$-M-D-H-M-L-T-S-T-R-S-$HA2_{(66-222)}$) [SEQ ID NO: 21 & 22], A ($NS1_{(1-81)}$-Q-I-P-$HA2_{(69-222)}$) [SEQ ID NO: 23 & 24], C ($NS1_{(1-81)}$-Q-I-P-$HA2_{(81-222)}$) [SEQ ID NO: 25 & 26], ΔD ($NS1_{(1-81)}HA2_{(150-222)}$) [SEQ ID NO: 27], Δ13 ($NS1_{(1-81)}$-D-L-S-R-$HA2_{(1-70)}$-S-C-L-T-A-Y-H-R) [SEQ ID NO: 28], M ($NS1_{(1-81)}$-Q-I-P-$HA2_{(65-196)}$-G-G-S-Y-S-M-E-H-F-R-W-G-K-P-V) [SEQ ID NO: 29], ΔM ($NS1_{(1-81)}$-Q-I-P-$HA2_{(65-196)}$-G-G-S-Y-S-M-L-V-N) [SEQ ID NO: 30], ΔM+ ($NS1_{(1-81)}$-Q-I-P-$HA2_{(65-200)}$-L-V-L-L) [SEQ ID NO: 31 & 32]. These H1N1 fusion proteins are described in published European Patent Application 366,238 and in copending U.S. Patent Application Ser. No. 07/751,896. Other suitable H1 proteins consist of unfused polypeptides, such as $H1HA2_{66-222}$ [SEQ ID NO: 33 & 34] which is disclosed in co-pending U. S. Patent Application Ser. No. 07/751,898, incorporated herein by reference. Thus, one desirable combination vaccine to provide protection against Type A

influenza contains $NS1_{(1-81)}H3HA2_{(1-221)}$ protein [SEQ ID NO: 9 & 10] of the invention, one or more proteins derived from subtype H1N1 as described above, and an aluminum adjuvant.

Preferably, a combination vaccine of the invention will contain an immunogenic amount of the H3 fusion protein of the invention in combination with immunogenic amounts of influenza antigens derived from the other type A influenza virus subtypes, including among others, H1, H2, H3, H4, H5, H6 and H7 as well as a type B fusion protein of the invention. Therefore, other preferred combination vaccines would include the $NS1_{(1-81)}H3HA2_{(77-221)}$ protein [SEQ ID NO: 11 & 12] in combination with one or more additional influenza antigens derived from the type or subtype influenza viruses described above. Thus, the combination vaccine will protect against influenza infections caused by both type A and type B influenza viruses. Still other combination vaccine compositions will employ other proteins described herein.

The compositions of the present invention are advantageously made up in a dose unit form adapted for the desired mode of administration. Each unit will contain, at a minimum, a predetermined quantity of the selected HA2 subunit protein, e.g., H3HA2 protein and/or

BHA2 protein, and adjuvant calculated to produce the desired therapeutic effect in optional association with a pharmaceutical diluent, carrier, or vehicle.

Dosage protocol can be optimized in accordance with standard vaccination practices. Typically, the vaccine will be administered intramuscularly, although other routes of administration may be used, such as intradermal. It is expected that an effective immunogenic amount of a protein, fusion protein or combination of proteins of this invention for average adult humans is in the range of 1 to 1000 micrograms. Another desirable immunogenic amount ranges between 50 to 500 micrograms. Most preferably, the proteins of the invention are in admixture with the same amount or more adjuvant to form a vaccine composition.

While the proteins described herein have been particularly developed for use in humans (e.g., the H3HA2 and BHA2 sequences), it is expected that due to species cross-reactivity, these vaccines will be useful in other animals, particularly swine. Additionally, similar molecules can be prepared for equine and avian veterinary applications utilizing the HA2 proteins from other strains to which animals are susceptible. Combination vaccines for use in swine would preferably include protections against both H1 and H3 viruses. Combination vaccines for use in equine would preferably include

protection against H3 and H7 viruses. Combination vaccines for use in avian species would preferably confer protection against H5 and H7 viruses. Appropriate dosages can be determined by one skilled in veterinary medicine.

It will be understood, however, that the specific effective immunogenic amount for any particular patient will depend upon a variety of factors including the age, general health, sex, and diet of the vaccinee; the species of the vaccinee; the time of administration; the route of administration; interactions with any other drugs being administered; and the degree of protection being sought.

The vaccine can be administered initially in late summer or early fall and can be readministered two to six weeks later, if desirable, or periodically as immunity wanes, for example, every two to five years. Of course, as stated above, the administration can be repeated at suitable intervals if necessary or desirable.

The following examples illustrate methods for preparing H3HA2 and BHA2 fusion proteins of the invention and demonstrate the subtype specific protection against heterologous virus induced upon vaccination with the H3HA2 proteins. These examples are illustrative only and do not limit the scope of the invention.

EXAMPLE 1 - PLASMID pMS3H3HA

Plasmid pFV88 contains the entire 221 amino acid length HA from A/Udorn, an H3 subtype virus [C. J. Lai et al, Proc. Natl. Acad. Sci. USA, 77:210-214 (1980)], which HA nucleic acid sequence is illustrated in Fig. 1 [SEQ ID NO: 1]. This plasmid was cut with Pst I. The resulting 1900 bp fragment, which contains the entire HA (HA1 and HA2) fragment and some GC tailing, was then inserted into pUC18 [Bethesda Research Laboratories]. The resulting plasmid is termed pMS3 or pMS3H3HA.

EXAMPLE 2 - pPMG1

Plasmid pAPR801 is a pBR322-derived cloning vector which carries the NS1 coding region (A/PR/8/34). It is described by Young et al, in The Origin of Pandemic Influenza Viruses, ed. by W. G. Laver, Elsevier Science Publishing Co. (1983).

Plasmid pAS1 is a pBR322-derived expression vector which contains the P_L promoter, an N utilization site (to relieve transcriptional polarity effects in the presence of N protein) and the cII ribosome binding site including the cII translation initiation codon followed immediately by a BamHI site. It is described by Rosenberg et al, in Methods Enzymol., 101:123-138 (1983).

Plasmid pAS1ΔEH was prepared by deleting a non-essential EcoRI-HindIII region of pBR322 origin from pAS1. A 1236 base pair BamHI fragment of pAPR801, containing the NS1 coding region in 861 base pairs of viral origin and 375 base pairs of pBR322 origin, was inserted into the BamHI site of pAS1ΔEH. The resulting plasmid, pAS1ΔEH/801 expresses authentic NS1 (230 amino acids). The plasmid has an NcoI site between the codons for amino acids 81 and 82 and an NruI site 3' to the NS sequences. The BamHI site between amino acids 1 and 2 is retained.

Plasmid pMG27N, a pAS1 derivative [Mol. Cell. Biol., 5:1015-1024 (1985)], was cut with BamHI and SacI and ligated to a BamHI/NcoI fragment encoding the first 81 amino acids of NS1 from pAS1ΔEH801 and a synthetic DNA NcoI/SacI fragment of the following sequence:

SEQ ID NO: 35:

5'-CATGGATCATATGTTAACAGATATCAAGGCCTGACTGACTGAGAGCT-3'

SEQ ID NO: 36:

3'- CTAGTATACAATTGTCTATAGTTCCGGACTGACTGACTC -5'

The resulting plasmid, pMG1, allows the insertion of DNA fragments after the first 81 amino acids of NS1 in any of the three reading frames within the synthetic linker fragment followed by termination codons in all three reading frames.

EXAMPLE 3 - pMG1H3HA

Plasmid pMG1, described above in Example 2, was digested with NcoI and XbaI, releasing a 54 bp fragment, which was discarded. pMS3H3HA, described in Example 1 above, was digested with HhaI and XbaI, and a 701 bp fragment containing the coding sequence for the HA2 subunit of influenza strain A/Udorn (H3N2) was isolated, as illustrated in Fig. 1 [SEQ ID NO: 1].

Synthetic oligonucleotides were annealed to generate an NcoI 5' overhang sequence (at the 5' end) and a HhaI 3' overhang sequence (at the 3' end). The sequence of these oligonucleotides is as follows:

SEQ ID NO: 37: 5'-CATGGGCGCCCATATGGGCATATTCGGCG-3'

SEQ ID NO: 38: 3'- CCGCGGGTATACCCGTATAAGCC -5'

The annealing reaction was performed as follows. The annealing mixture was made up of 2.5μL each of 5' oligo (1.3 μg/μL), the 3' oligo (1.2 μg/μL), and added water (15 μL) to a final volume of 20 μL. The reaction tubes were then placed in 4 mL culture tubes containing water which had been heated to 65°C for 10 minutes and allowed to cool down slowly. The tubes were then put on ice and used immediately for ligation.

This three part ligation generates pMG1H3HA2$_{(1-221)}$ [SEQ ID NO: 9] which codes for the first 81 amino acids of NS1 fused to four amino acids donated from the linker and amino acids 1-221 of the HA2 subunit. This sequence

is illustrated in Fig. 2 [SEQ ID NO: 9 & 10]. This molecule is also designated $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 9 & 10].

EXAMPLE 4 - $NS1_{(1-81)}H3HA2_{(77-221)}$ [SEQ ID NO: 11 & 12]

pMS3H3HA, described in Example 1 above, was digested with EcoRI and end-filled (Klenow). Subsequently, the vector was digested with XbaI. A 487 bp fragment, which contains the coding sequence for amino acids 77-221 of the HA2 subunit, was isolated and ligated to the HpaI and XbaI sites of pMG1. The resulting vector codes for a fusion polypeptide containing amino acids 1-81 of NS1 fused to amino acids 77-221 of the HA2 subunit. This molecule has been termed $NS1_{(1-81)}H3HA2_{(77-221)}$ and is illustrated in Fig. 3 [SEQ ID NO: 11 & 12].

EXAMPLE 5 - $pMG_{42}BLHA2$

To derive a vector similar to pMG1 (described in Example 2), which contains the coding region for the first 42 amino acids of NS1 rather than the first 81 amino acids of NS1, pMG1 was digested with BamHI and NcoI and ligated to the BamHI/NcoI fragment encoding amino acids 2 to 42 of NS1 from $pNS1_{42}TGF\alpha$. $pNS1_{42}TGF\alpha$ is derived when pAS1ΔEH801 is cut with NcoI and SalI and ligated to a synthetic DNA encoding human TGFα as an

NcoI/SalI fragment. $pNS1_{42}TGF\alpha$ encodes a protein comprised of the first 42 amino acids of NS1 and the mature TGFα sequence. The NS1 portion of $pNS1_{42}TGF\alpha$ contains an amino acid change from Cys to Ser at amino acid #13.

The resulting plasmid, termed $pMG_{42}A$, was then modified to contain an alternative synthetic linker after the $NS1_{42}$ sequence with a different set of restriction enzyme sites within which to insert foreign DNA fragments into the three reading frames after the $NS1_{42}$. This linker has the following sequence:

SEQ ID NO: 39:

5'-CATGGATCATATGTTAACAAGTACTCGATATCAATGAGTGACTGAAGCT-3'

SEQ ID NO: 40:

3'- CTAGTATACAATTGTTCATGAGCTATAGTTACTCACTGACT -5'

The resulting plasmid is called $pMG_{42}B$. This vector is needed to contain the neomycin phosphotransferase-1 (NPT-1) gene which confers kanamycin resistance.

As described in Shatzman and Rosenberg, Met. Enzymol., 152:661-673 (1987), pOTS207 is a pAS derived cloning vector which carries the kanamycin resistance gene from Tn903 [Berg et al, Microbiology, ed. D. Schlessinger, pp. 13-15, American Society for Microbiology (Washington, DC 1978); Nomura et al, The Single-Stranded DNA Phages, ed. D. Denhardt et al,

pp.467-472, Cold Spring Harbor Laboratory (New York 1978); Castellazzi et al, Molecul. Gen. Genet., 117:211-218 (1982)]. It was constructed by digesting plasmid pUC8 [Yanisch-Perron et al, Gene, 33:103-119 (1985)], with BamHI and ligated to a BcII fragment containing the kanamycin gene from Tn903. The resulting plasmid, pUC8-Kan, was digested with EcoRI and PstI, and the fragment containing the kanamycin gene was inserted between the EcoRI and PstI sites of pOTSV [Shatzman and Rosenberg, cited above]. The resulting plasmid is pOTS207.

The pOTS207 was digested with EcoRI and PstI, and the 1467 bp fragment containing the kanamycin resistance gene was isolated. Synthetic oligonucleotides:

SEQ ID NO: 41: 5' AATTCGTACCTA 3'

SEQ ID NO: 42: 3' GCATGGATCTAG 5'

were made to link the NPT-1 gene to pMG42B vector. $pMG_{42}B$ was digested with BglII and PstI. The EcoRI/PstI NPT-1 gene fragment and the synthetic oligo linker were ligated to the digested $pMG_{42}B$. The resulting plasmid, $pMG_{42}Kn$ allows fusions, in three different reading frames, to the NS_{1-42} gene, while allowing antibiotic selection with kanamycin.

Plasmid pBHA is a pBR322-derived vector, containing the complete nucleotide sequence of the hemagglutinin (HA) gene of a type B influenza virus

(B/Lee/40). It is described by Krystal et al, Proc. Natl. Acad. Sci. USA, 79:4900-4804 (1982). pBHA was digested with RsaI and a 813 bp fragment containing the HA subunit was isolated. This fragment was ligated into plasmid $pMG_{42}Kn$ (described above) that had been digested with ScaI. During the cloning, a base (T) was deleted from the ScaI recognition site shifting the gene out of the reading frame. The vector was digested with NcoI, and filled-in using Klenow, putting the gene back into the reading frame.

The resulting construct, $pMG_{42}BLHA2$ [SEQ ID NO: 14], expresses a fusion polypeptide containing amino acids 1-42 of NS1 and 41-233 of the HA2 subunit. This construct contains the Cys to Ser change at amino acid #13 of the NS1 portion of the fusion peptide.

In preliminary studies with this construct, vaccinated laboratory mice demonstrated protection from challenge with type B influenza in the absence of neutralizing antibody for the virus.

EXAMPLE 6 - PREPARING SEED VIRUS AND RAISING ANTISERA

The seed virus, A/Udorn, was prepared according to the procedures described in P. Palese and J. Schulman, Virol., 57:227-237 (1974). Briefly, this technique is as follows.

Influenza virus strain A/Udorn was inoculated in 10-day old embryonated hen's eggs into the allantoic cavity. The eggs were incubated for 24-48 hours at 35°C then chilled at 4°C overnight. A portion of the eggshell over the airsac was removed and the allantoic fluid was aseptically removed using a 10-ml syringe. The fluid was centrifuged at low speed (3,000 x g) to remove particulates. This clarified supernatant was centrifuged at high speed using an SW28 Beckman rotor at 27,000 rpm (4°C for 90 minutes), resulting in the virus pellet. The virus was resuspended in 10 mM Tris (pH 7.5) containing 100 mM NaCl, 1 mM EDTA and repelleted as before. The virus was layered on 30-60% sucrose gradient in 1 mM EDTA (NTE) and spun for 3-5 hours at 25,000 rpm. The band in the middle of the tube was withdrawn, diluted in NTE and centrifuged at 27,000 rpm for 90 minutes. The pellet was suspended in phosphate-buffered saline (PBS). These viral particles were used as immunogens for preparation of antisera.

Antisera was prepared as follows. 100-200 micrograms of purified virus in complete Freund's adjuvant was injected into the subscapula of a New Zealand White rabbit. A second injection in incomplete Freund's adjuvant was done 4 weeks later, and the animals were bled 7-10 days later.

EXAMPLE 7 - EXPRESSION OF H3HA2 FUSION PROTEINS

A. $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 9 & 10]

The plasmid $pMG1H3HA2_{(1-221)}$ [SEQ ID NO: 9] was transfected into E. coli strain AR58 [SmithKline Beecham Pharmaceuticals]. Cultures were grown at 32°C to mid-log phase at which time cultures were shifted to 39.5°C for 2 hours. The E. coli cell pellets containing the recombinant polypeptide were then stored at -70°C until used.

Production of the $NS1_{(1-81)}H3HA2_{(1-221)}$ protein [SEQ ID NO: 10] was confirmed by Western blot analysis [Towbin et al, Proc. Natl. Acad. Sci. U.S.A., 76:4350 (1979)] using antisera prepared against A/Udorn virus, as described in Example 5. A major immunoreactive species was found at a molecular weight of 35,050 daltons.

B. $NS1_{(1-81)}H3HA2_{(77-221)}$ [SEQ ID NO: 11 & 12]

The plasmid encoding the $NS1_{(1-81)}H3HA2_{(77-221)}$ peptide [SEQ ID NO: 11 & 12] was expressed as described in part A above. Production of this peptide was confirmed by Western blot analysis, as described above. A major immunoreactive species was found at a molecular weight of 26,697 daltons.

EXAMPLE 8 - PARTIAL PURIFICATION OF H3HA2 FUSION PROTEINS

E. coli cell pellets containing the recombinant polypeptides, prepared as described in Example 6, were stored at -70°C until used. E. coli cells were thawed and resuspended in lysis buffer A (50 mM Tris-HCl, 5% glycerol, 2 mM EDTA and 0.1 mM DTT, pH 8.0) at 10 mL/gram. The stirred suspension was then treated with lysozyme (0.2 mg/mL) for 45 minutes at room temperature and sonicated 2x for 2-3 minutes each time by a Sonicator. The resultant suspension was treated with 0.1% DOC for 60 minutes at 4°C, then centrifuged at 25,000 x g. The pellet was resuspended by sonication in 50 mM glycine pH 10.0, 5% glycerol, 2 mM EDTA and then the suspension was treated with 1% Triton X-100 [J.T. Baker Chemicals Co.] at 4°C for 60 minutes and centrifuged as above.

The resulting pellet was solubilized in 50 mM Tris, 8 M urea, pH 8.0 and centrifuged to remove any insoluble material. This solubilized material is dialyzed against 10 mM Tris, 1 mM EDTA, pH 8.0 followed, again, by centrifugation of insoluble material. The solubilized material is designated as "crude" material and is used in in vitro and in vivo mouse assays. At this point, the material is approximately 40 - 50% pure.

The "crude" material was electrophoresed through an SDS-PAGE and the appropriate H3HA2 protein bands were visualized by KCl staining according to D. Hager et al, Anal. Biochem, 109:76-86 (1980). The band was cut-out and eluted electrophoretically by the "S&S Elutrap Electro-Separation System" [Schleicher & Schuell]. The electro-eluting buffer was the Tris-glycine. A concentrated and eluted sample was obtained and exhaustively dialyzed against 0.01 M NH_4HCO_3 and 0.02% SDS [M. Hunkapiller et al, Method. Enzymol., 91:227-236 (1983)]. This sample was frozen quickly by dry ice and lyophilized to complete dryness. The lyophilized material was brought back into solution using 50 mM Tris pH 8.0 and used for in vitro and in vivo mouse assays.

Following this gel elution step, the protein is usually greater than 75% pure.

EXAMPLE 9 - H3 SUBTYPE HETEROLOGOUS PROTECTION ELICITED BY VACCINATION WITH $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10]

Mice (NIH/Swiss; 15 per group) were vaccinated subcutaneously with 50 or 10 μg $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 9 & 10] in aluminum hydroxide on days 0 and 21. The mice were boosted intraperitoneally on day 42 with the protein without adjuvant. On day 47, mice were challenged intranasally with 2 - 3 LD_{50} doses of either A/PR/8/34 (H1N1) or A/HK/68 (H3N2) virus, and survival was

monitored through day 21. This represents a heterologous challenge (A/PR/8/34) and an H3 heterosubtypic challenge, since the $NS1_{(1-81)}H3HA2_{(1-221)}$ construct [SEQ ID NO: 9 & 10] was derived from A/Udorn/72 cDNA. The control group received adjuvant (CFA) only.

The results in Table 1 below show that survival in mice vaccinated with $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10] and challenged with A/HK/68 (80-93%) was significantly higher than in control mice which were injected with adjuvant only (26% survival). In contrast, vaccination with $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10] did not confer protection against challenge with A/PR/8/34, an H1N1 strain (0-26% survival). Thus protection elicited by $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10] is selective for antigenically diverse virus strains within the H3 subtype.

Likewise, vaccination with the D protein ($NS1_{(1-81)}HA2_{(65-222)}$ [SEQ ID NO: 18], derived from the H1N1 subtype) elicits protection from heterosubtypic challenge with H1N1, but not the H3N2 subtype [S Dillon et al, Nature, in press (1992); Mbawuike et al, Faseb. J., 5:A1362 (abs. 5749 and Table 1]. These results in outbred mice also suggest that the response to the H1 and H3 proteins will not be restricted to a limited number of individuals with certain major histocompatibility alleles, and therefore the vaccine will be effective in a majority of individuals.

Table 1

Percent Survival After Challenge:

Immunization	HA Subtype	A/PR/8/34 (H1N1)	A/HK/68 (H3N2)
50 μg $NS1_{1-81}H3HA2_{1-221}$	H3	26	80*
10 μg $NS1_{1-81}H3HA2_{1-221}$	H3	0	93*
10 μg $NS1_{1-81}HA2_{1-221}$	H1	67*	13
A/HK/68 virus	H3	60*	100*
Control (Al^{+3})	-	0	26

$p \leq 0.05$ vs. control in Fishers exact probability test

Vaccination of mice with live homologous (A/HK/68) virus provided complete or partial protection, reflecting protection mediated by neutralizing antibody (homologous H3N2 challenge) and/or CTL (heterologous H1N1 challenge), respectively.

Duration of protective immunity was tested by immunizing mice subcutaneously with the recombinant influenza protein plus adjuvant on days 0 and 21. Some mice were also given an ip injection of the protein (without adjuvant) on day 42. Mice were challenged with A/HK/68 (H3N2) on day 47, four weeks after the second injection. Control mice were immunized as described above for Table 1, where an ip injection was given at week 6 (5 days prior to challenge). The results in Table 2 show that $CB6F_1$ mice (15 per group) were significantly protected when challenged with the A/HK/68 heterologous H3 virus strain 5-28 days after the last injection.

Table 2

Dose (μg per injection) of $NS1_{1-81}H3HA2_{1-221}$	Adjuvant	Injection Schedule	Percent Survival
50 μg	CFA	0,21	86*
50 μg	CFA	0,21,42	100*
0 μg	CFA	0,21	6
50 μg	Al^{+3}	0,21	93*
50 μg	Al^{+3}	0,21,42	93*
0 μg	Al^{+3}	0,21	0

*$p \leq 0.05$ v. control in Fisher's exact probability test

EXAMPLE 10 - TYPE A CROSS-PROTECTION WITH D AND H3C13 PROTEIN

Mice ($CB6F_1$) were divided randomly into six groups, with fifteen in each group. The mice were injected subcutaneously with proteins in Al^{+3} (100 μg) on days 0 and 21, and then were challenged with 2-3 LD_{50} doses of virus on day 49. Survival was monitored through day 21. The results of this study are illustrated in Table 3 below. For convenience, $NS1_{1-81}H3HA2_{1-221}$ is referred to as H3C13 in the table below.

Table 3

	Immunization	HA Subtype	Percent Survival After Challenge with: A/PR/8/34 (H1N1)	A/HK/68 (H3N2)
1.	50 μg H3C13 50 μg D	H3 H1	73*	73*
2.	10 μg H3C13 10 μg D	H3 H1	67*	100*
3.	1 μg H3C13 1 μg D	H3 H1	86*	73*
4.	50 μg H3C13	H3	7	73*
5.	50 μg D	H1	47**	7
6.	Al^{+3} control	-	7	0

* $p \leq 0.001$ vs. control group
** $p \leq 0.03$ vs. control group

This data demonstrates that mice immunized with a mixture of the D protein and H3C13 protein in aluminum adjuvant were protected against challenge with either A/PR/8/34 (H1) or A/HK/68 (H3) virus. In contrast, mice immunized with the D protein were protected against H1 but not H3 challenge. Likewise, mice immunized with the H3C13 protein were protected against the H3 but not the H1 challenge. Therefore, the combination of the D protein and the H3C13 proteins elicited protection against the currently circulating subtypes of influenza A virus. Thus, this combination represents a subtype cross-protective vaccine.

Numerous modifications and variations of the present invention are included in the above-identified specification and are expected to be obvious to one of skill in the art. Such modifications and alterations to the compositions and processes of the present invention are believed to be encompassed in the scope of the claims appended hereto.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Shatzman, Allan
Scott, Miller
Dillon, Susan B.

(ii) TITLE OF INVENTION: Vaccinal Polypeptides

(iii) NUMBER OF SEQUENCES: 42

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SmithKline Beecham Corporation - Corporate Patents
(B) STREET: U.S. Mailcode VW2220 - 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: Pennsylvania
(E) COUNTRY: USA
(F) ZIP: 19406-2799

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Canter, Carol G.
(B) REGISTRATION NUMBER: 31,151
(C) REFERENCE/DOCKET NUMBER: SBC14224-8

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 215-270-5013
(B) TELEFAX: 215-270-5090

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 666 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..663

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA AAT GGT TGG GAG GGA      48
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
 1               5                  10                  15

ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT CAA AAT TCT GAG GGC ACA      96
Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
            20                  25                  30

GGA CAA GCA GCA GAT CTT AAA AGC ACT CAA GCA GCC ATC GAC CAA ATC     144
Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
        35                  40                  45

AAT GGG AAA CTG AAT AGG GTA ATC GAG AAG ACG AAC GAG AAA TTC CAT     192
Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
    50                  55                  60

CAA ATC GAA AAG GAA TTC TCA GAA GTA GAA GGG AGA ATT CAG GAC CTC     240
Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65                  70                  75                  80

GAG AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC TGG TCT TAC AAT GCG     288
Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                85                  90                  95

GAG CTT CTT GTC GCT CTG GAG AAC CAA CAT ACA ATT GAT CTG ACT GAC     336
Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

TCG GAA ATG AAC AAA CTG TTT GAA AAA ACA AGG AGG CAA CTG AGG GAA     384
Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

AAT GCT GAG GAC ATG GGC AAT GGT TGC TTC AAA ATA TAC CAC AAA TGT     432
Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG ACT TAT GAC CAT GAT     480
Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

GTA TAC AGA GAC GAA GCA TTA AAC AAC CGG TTT CAG ATC AAA GGT GTT     528
Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

GAA CTG AAG TCA GGA TAC AAA GAC TGG ATC CTG TGG ATT TCC TTT GCC     576
Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
            180                 185                 190

ATA TCA TGC TTT TTG CTT TGT GTT GTT TTG CTG GGG TTC ATC ATG TGG     624
Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
        195                 200                 205

GCC TGC CAG AAA GGC AAC ATT AGG TGC AAC ATT TGC ATT TGA             666
Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
 1               5                  10                  15

Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
            20                  25                  30

Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
        35                  40                  45

Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
    50                  55                  60

Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65                  70                  75                  80

Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                85                  90                  95

Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
            180                 185                 190

Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
        195                 200                 205

Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 666 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..663

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA AAT GGT TGG GAG GGA          48
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
 1               5                  10                  15

ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT CAA AAT TCC GAG GGC ACA          96
Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
            20                  25                  30

GGA CAA GCA GCA GAT CTT AAA AGC ACT CAA GCA GCC ATC GAC CAA ATC         144
Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
        35                  40                  45

AAT GGG AAA CTG AAT AGG GTA ATC GAG AAG ACG AAC GAG AAA TTC CAT         192
Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
    50                  55                  60

CAA ATC GAA AAG GAA TTC TCA GAA GTA GAA GGG AGA ATT CAG GAC CTC         240
Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65                  70                  75                  80

GAG AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC TGG TCT TAC AAT GCG         288
Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                85                  90                  95

GAG CTT CTT GTC GCT CTG GAG AAC CAA CAT ACA ATT GAT CTG ACT GAC         336
Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

TCG GAA ATG AAC AAA CTG TTT GAA AAA ACA AGG AGG CAA CTG AGG GAA         384
Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

AAT GCT GAG GAC ATG GGC AAT GGT TGC TTC AAA ATA TAC CAC AAA TGT         432
Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG ACT TAT GAC CAT GAT         480
Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

GTA TAC AGA GAC GAA GCA TTA AAC AAC CGG TTT CAG ATC AAA GGT GTT         528
Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

GAA CTG AAG TCA GGA TAC AAA GAC TGG ATC CTG TGG ATT TCC TTT GCC         576
Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
            180                 185                 190
```

```
ATA TCA TGC TTT TTG CTT TGT GTT GTT TTG CTG GGG TTC ATC ATG TGG        624
Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
        195                 200                 205

GCC TGC CAA AAA GGC AAC ATT AGG TGC AAC ATT TGC ATT TGA               666
Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
 1               5                  10                  15

Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
            20                  25                  30

Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
        35                  40                  45

Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
    50                  55                  60

Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65                  70                  75                  80

Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                85                  90                  95

Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
            180                 185                 190

Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
        195                 200                 205

Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 670 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..666

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
GGT CTA TTT GGA GCC ATT GCC GGT TTT ATT GAA GGG GGA TGG ACT GGA        48
Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly
 1               5                  10                  15

ATG ATA GAT GGA TGG TAC GGT TAT CAT CAT CAG AAT GAA CAG GGA TCA        96
Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn Glu Gln Gly Ser
            20                  25                  30

GGC TAT GCA GCG GAT CAA AAA AGC ACA CAA AAT GCC ATT AAC GGG ATT       144
Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala Ile Asn Gly Ile
        35                  40                  45

ACA AAC AAG GTG AAC TCT GTT ATC GAG AAA ATG AAC ATT CAA TTC ACA       192
Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn Ile Gln Phe Thr
    50                  55                  60

GCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG ATG GAA AAT TTA       240
Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu
65                  70                  75                  80

AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA TAT AAT GCA       288
Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala
                85                  90                  95

GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT TTC CAT GAC       336
Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp
            100                 105                 110

TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA TTA AAG AAT       384
Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn
        115                 120                 125

AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC CAC AAG TGT       432
Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys
    130                 135                 140

GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT GAT TAT CCC       480
Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
145                 150                 155                 160

AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA GAT GGA GTG       528
Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
                165                 170                 175
```

```
AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC TAC TCA ACT      576
Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
            180                 185                 190

GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA ATC AGT TTC      624
Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
        195                 200                 205

TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC ATC              666
Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
    210                 215                 220

TGAG                                                                670
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 222 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly
 1               5                  10                  15

Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn Glu Gln Gly Ser
            20                  25                  30

Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala Ile Asn Gly Ile
        35                  40                  45

Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn Ile Gln Phe Thr
    50                  55                  60

Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu
65                  70                  75                  80

Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala
                85                  90                  95

Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp
            100                 105                 110

Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn
        115                 120                 125

Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys
    130                 135                 140

Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
145                 150                 155                 160

Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
                165                 170                 175
```

Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
180 185 190

Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
195 200 205

Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
210 215 220

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 670 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..670

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGCATATTCG GCGCAATAGC AGGTTTCATA GAAAATGGTT GGGAGGGAAT GATAGACGGT 60

TGGTACGGTT TCAGGCATCA AAATTCNGAG GGCACAGGAC AAGCAGCAGA TCTTAAAAGC 120

ACTCAAGCAG CCATCGACCA AATCAATGGG AAACTGAATA GGGTAATCGA GAAGACGAAC 180

GAGAAATTCC ATCAAATCGA AAAGGAATTC TCAGAAGTAG AAGGGAGAAT TCAGGACCTC 240

GAGAAATACG TTGAAGACAC TAAAATAGAT CTCTGGTCTT ACAATGCGGA GCTTCTTGTC 300

GCTCTGGAGA ACCAACATAC AATTGATCTG ACTGACTCGG AAATGAACAA ACTGTTTGAA 360

AAAACAAGGA GGCAACTGAG GGAAAATGCT GAGGACATGG GCAATGGTTG CTTCAAAATA 420

TACCACAAAT GTGACAATGC TTGCATAGGG TCAATCAGAA ATGGGACTTA TGACCATGAT 480

GTATACAGAG ACGAAGCATT AAACAACCGG TTTCAGATCA AAGGTGTTGA ACTGAAGTCA 540

GGATACAAAG ACTGGATCCT GTGGATTTCC TTTGCCATAT CATGCTTTTT GCTTTGTGTT 600

GTTTTGCTGG GGTTCATCAN NNTGTGGGCC TGCCANAAAG GCAACATTAG GTGCAACATT 660

TGCATTTGAN 670

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 222 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
1               5                  10                  15

Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
            20                  25                  30

Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
        35                  40                  45

Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
    50                  55                  60

Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65                  70                  75                  80

Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                85                  90                  95

Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

Glu Leu Lys Ser Xaa Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe
            180                 185                 190

Ala Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met
        195                 200                 205

Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 918 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..918

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG        48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC        96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC       144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA       192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC       240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

ATG GGC GCC CAT ATG GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA       288
Met Gly Ala His Met Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu
                85                  90                  95

AAT GGT TGG GAG GGA ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT CAA       336
Asn Gly Trp Glu Gly Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln
            100                 105                 110

AAT TCT GAG GGC ACA GGA CAA GCA GCA GAT CTT AAA AGC ACT CAA GCA       384
Asn Ser Glu Gly Thr Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala
        115                 120                 125

GCC ATC GAC CAA ATC AAT GGG AAA CTG AAT AGG GTA ATC GAG AAG ACG       432
Ala Ile Asp Gln Ile Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr
    130                 135                 140

AAC GAG AAA TTC CAT CAA ATC GAA AAG GAA TTC TCA GAA GTA GAA GGG       480
Asn Glu Lys Phe His Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly
145                 150                 155                 160

AGA ATT CAG GAC CTC GAG AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC       528
Arg Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu
                165                 170                 175

TGG TCT TAC AAT GCG GAG CTT CTT GTC GCT CTG GAG AAC CAA CAT ACA       576
Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr
            180                 185                 190

ATT GAT CTG ACT GAC TCG GAA ATG AAC AAA CTG TTT GAA AAA ACA AGG       624
Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg
        195                 200                 205

AGG CAA CTG AGG GAA AAT GCT GAG GAC ATG GGC AAT GGT TGC TTC AAA       672
Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys
    210                 215                 220

ATA TAC CAC AAA TGT GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG       720
Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly
225                 230                 235                 240
```

```
ACT TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA AAC AAC CGG TTT        768
Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe
               245                 250                 255

CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC AAA GAC TGG ATC CTG        816
Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu
           260                 265                 270

TGG ATT TCC TTT GCC ATA TCA TGC TTT TTG CTT TGT GTT GTT TTG CTG        864
Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu
       275                 280                 285

GGG TTC ATC ATG TGG GCC TGC CAA AAA GGC AAC ATT AGG TGC AAC ATT        912
Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile
   290                 295                 300

TGC ATT                                                               918
Cys Ile
305
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 306 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gly Ala His Met Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu
                85                  90                  95

Asn Gly Trp Glu Gly Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln
            100                 105                 110

Asn Ser Glu Gly Thr Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala
        115                 120                 125

Ala Ile Asp Gln Ile Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr
    130                 135                 140

Asn Glu Lys Phe His Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly
145                 150                 155                 160
```

```
Arg Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu
                165                 170                 175

Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr
            180                 185                 190

Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg
        195                 200                 205

Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys
    210                 215                 220

Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly
225                 230                 235                 240

Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe
                245                 250                 255

Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu
            260                 265                 270

Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu
        275                 280                 285

Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile
    290                 295                 300

Cys Ile
305
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..690

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG      48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC      96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC     144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA     192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60
```

```
GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC      240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

ATG GAT CAT ATG TTA ATT CAG GAC CTC GAG AAA TAC GTT GAA GAC ACT      288
Met Asp His Met Leu Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr
                 85                  90                  95

AAA ATA GAT CTC TGG TCT TAC AAT GCG GAG CTT CTT GTC GCT CTG GAG      336
Lys Ile Asp Leu Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu
            100                 105                 110

AAC CAA CAT ACA ATT GAT CTG ACT GAC TCG GAA ATG AAC AAA CTG TTT      384
Asn Gln His Thr Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe
        115                 120                 125

GAA AAA ACA AGG AGG CAA CTG AGG GAA AAT GCT GAG GAC ATG GGC AAT      432
Glu Lys Thr Arg Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn
    130                 135                 140

GGT TGC TTC AAA ATA TAC CAC AAA TGT GAC AAT GCT TGC ATA GGG TCA      480
Gly Cys Phe Lys Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser
145                 150                 155                 160

ATC AGA AAT GGG ACT TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA      528
Ile Arg Asn Gly Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu
                165                 170                 175

AAC AAC CGG TTT CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC AAA      576
Asn Asn Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys
            180                 185                 190

GAC TGG ATC CTG TGG ATT TCC TTT GCC ATA TCA TGC TTT TTG CTT TGT      624
Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys
        195                 200                 205

GTT GTT TTG CTG GGG TTC ATC ATG TGG GCC TGC CAA AAA GGC AAC ATT      672
Val Val Leu Leu Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile
    210                 215                 220

AGG TGC AAC ATT TGC ATT                                             690
Arg Cys Asn Ile Cys Ile
225                 230
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 230 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30
```

```
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Asp His Met Leu Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr
                85                  90                  95

Lys Ile Asp Leu Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu
            100                 105                 110

Asn Gln His Thr Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe
        115                 120                 125

Glu Lys Thr Arg Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn
    130                 135                 140

Gly Cys Phe Lys Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser
145                 150                 155                 160

Ile Arg Asn Gly Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu
                165                 170                 175

Asn Asn Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys
            180                 185                 190

Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys
        195                 200                 205

Val Val Leu Leu Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile
    210                 215                 220

Arg Cys Asn Ile Cys Ile
225                 230
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 699 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..699

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TCC TTT CTT TGG          48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Ser Phe Leu Trp
 1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC          96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC ATG CAT GGA TCA TAT GTT         144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met His Gly Ser Tyr Val
        35                  40                  45

AAC AAG ACA CAA GAA GCT ATA AAC AAG ATA ACA AAA AAT CTC AAC TAT         192
Asn Lys Thr Gln Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Tyr
    50                  55                  60

TTA AGT GAG CTA GAA GTA AAA AAC CTT CAA AGA CTA AGC GGA GCA ATG         240
Leu Ser Glu Leu Glu Val Lys Asn Leu Gln Arg Leu Ser Gly Ala Met
65                  70                  75                  80

AAT GAG CTT CAC GAC GAA ATA CTC GAG CTA GAC GAA AAA GTG GAT GAT         288
Asn Glu Leu His Asp Glu Ile Leu Glu Leu Asp Glu Lys Val Asp Asp
                85                  90                  95

CTA AGA GCT GAT ACA ATA AGC TCA CAA ATA GAG CTT GCA GTC TTG CTT         336
Leu Arg Ala Asp Thr Ile Ser Ser Gln Ile Glu Leu Ala Val Leu Leu
            100                 105                 110

TCC AAC GAA GGG ATA ATA AAC AGT GAA GAT GAG CAT CTC TTG GCA CTT         384
Ser Asn Glu Gly Ile Ile Asn Ser Glu Asp Glu His Leu Leu Ala Leu
        115                 120                 125

GAA AGA AAA CTG AAG AAA ATG CTT GGC CCC TCT GCT GTA GAA ATA GGG         432
Glu Arg Lys Leu Lys Lys Met Leu Gly Pro Ser Ala Val Glu Ile Gly
    130                 135                 140

AAT GGG TGC TTT GAA ACC AAA CAC AAA TGC AAC CAG ACT TGC CTA GAC         480
Asn Gly Cys Phe Glu Thr Lys His Lys Cys Asn Gln Thr Cys Leu Asp
145                 150                 155                 160

AGG ATA GCT GCT GGC ACC TTT AAT GCA GGA GAT TTT TCT CTT CCC ACT         528
Arg Ile Ala Ala Gly Thr Phe Asn Ala Gly Asp Phe Ser Leu Pro Thr
                165                 170                 175

TTT GAT TCA TTA AAC ATT ACT GCT GCA TCT TTA AAT GAT GAT GGC TTG         576
Phe Asp Ser Leu Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu
            180                 185                 190

GAT AAT CAT ACT ATA CTG CTC TAC TAC TCA ACT GCT GCT TCT AGC TTG         624
Asp Asn His Thr Ile Leu Leu Tyr Tyr Ser Thr Ala Ala Ser Ser Leu
        195                 200                 205

GCT GTA ACA TTA ATG ATA GCT ATC TTC ATT GTC TAC ATG GTC TCC AGA         672
Ala Val Thr Leu Met Ile Ala Ile Phe Ile Val Tyr Met Val Ser Arg
    210                 215                 220

GAC AAT GTT TCT TGT TCC ATC TGT CTG                                    699
Asp Asn Val Ser Cys Ser Ile Cys Leu
225                 230
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 233 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Ser Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met His Gly Ser Tyr Val
        35                  40                  45

Asn Lys Thr Gln Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Tyr
    50                  55                  60

Leu Ser Glu Leu Glu Val Lys Asn Leu Gln Arg Leu Ser Gly Ala Met
65                  70                  75                  80

Asn Glu Leu His Asp Glu Ile Leu Glu Leu Asp Glu Lys Val Asp Asp
                85                  90                  95

Leu Arg Ala Asp Thr Ile Ser Ser Gln Ile Glu Leu Ala Val Leu Leu
            100                 105                 110

Ser Asn Glu Gly Ile Ile Asn Ser Glu Asp Glu His Leu Leu Ala Leu
        115                 120                 125

Glu Arg Lys Leu Lys Lys Met Leu Gly Pro Ser Ala Val Glu Ile Gly
    130                 135                 140

Asn Gly Cys Phe Glu Thr Lys His Lys Cys Asn Gln Thr Cys Leu Asp
145                 150                 155                 160

Arg Ile Ala Ala Gly Thr Phe Asn Ala Gly Asp Phe Ser Leu Pro Thr
                165                 170                 175

Phe Asp Ser Leu Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu
            180                 185                 190

Asp Asn His Thr Ile Leu Leu Tyr Tyr Ser Thr Ala Ala Ser Ser Leu
        195                 200                 205

Ala Val Thr Leu Met Ile Ala Ile Phe Ile Val Tyr Met Val Ser Arg
    210                 215                 220

Asp Asn Val Ser Cys Ser Ile Cys Leu
225                 230
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 924 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..921

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG          48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC          96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC         144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA         192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC         240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

ATG GAT CTG TCC AGA GGT CTA TTT GGA GCC ATT GCC GGT TTT ATT GAA         288
Met Asp Leu Ser Arg Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu
                85                  90                  95

GGG GGA TGG ACT GGA ATG ATA GAT GGA TGG TAC GGT TAT CAT CAT CAG         336
Gly Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln
            100                 105                 110

AAT GAA CAG GGA TCA GGC TAT GCA GCG GAT CAA AAA AGC ACA CAA AAT         384
Asn Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn
        115                 120                 125

GCC ATT AAC GGG ATT ACA AAC AAG GTG AAC TCT GTT ATC GAG AAA ATG         432
Ala Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met
    130                 135                 140

AAC ATT CAA TTC ACA GCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA         480
Asn Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys
145                 150                 155                 160

AGG ATG GAA AAT TTA AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT         528
Arg Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile
                165                 170                 175
```

```
TGG ACA TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT      576
Trp Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr
            180                 185                 190

CTG GAT TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA      624
Leu Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys
        195                 200                 205

AGC CAA TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG      672
Ser Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu
    210                 215                 220

TTC TAC CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG      720
Phe Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly
225                 230                 235                 240

ACT TAT GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA      768
Thr Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu
                245                 250                 255

AAG GTA GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG      816
Lys Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu
            260                 265                 270

GCG ATC TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG      864
Ala Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu
        275                 280                 285

GGG GCA ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA      912
Gly Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg
    290                 295                 300

ATA TGC ATC TGA                                                     924
Ile Cys Ile
305
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 307 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80
```

```
Met Asp Leu Ser Arg Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu
                85                  90                  95

Gly Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln
            100                 105                 110

Asn Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn
        115                 120                 125

Ala Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met
    130                 135                 140

Asn Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys
145                 150                 155                 160

Arg Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile
                165                 170                 175

Trp Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr
            180                 185                 190

Leu Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys
        195                 200                 205

Ser Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu
    210                 215                 220

Phe Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly
225                 230                 235                 240

Thr Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu
                245                 250                 255

Lys Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu
            260                 265                 270

Ala Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu
        275                 280                 285

Gly Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg
    290                 295                 300

Ile Cys Ile
305
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 729 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..726

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG          48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC          96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC         144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA         192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC         240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

ATG CAG ATC CCG GCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG         288
Met Gln Ile Pro Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
                85                  90                  95

ATG GAA AAT TTA AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG         336
Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
            100                 105                 110

ACA TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG         384
Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
        115                 120                 125

GAT TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC         432
Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
    130                 135                 140

CAA TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC         480
Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
145                 150                 155                 160

TAC CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT         528
Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
                165                 170                 175

TAT GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG         576
Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
            180                 185                 190

GTA GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG         624
Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
        195                 200                 205

ATC TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG         672
Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
    210                 215                 220
```

```
GCA ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA      720
Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
225                 230                 235                 240

TGC ATC TGA                                                      729
Cys Ile
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 242 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
                85                  90                  95

Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
            100                 105                 110

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
        115                 120                 125

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
    130                 135                 140

Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
145                 150                 155                 160

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
                165                 170                 175

Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
            180                 185                 190

Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
        195                 200                 205

Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
    210                 215                 220
```

```
Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
225                 230                 235                 240

Cys Ile
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 810 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..807

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG        48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC        96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC ATG GAT CTG TCC AGA GGT       144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met Asp Leu Ser Arg Gly
        35                  40                  45

CTA TTT GGA GCC ATT GCC GGT TTT ATT GAA GGG GGA TGG ACT GGA ATG       192
Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly Met
    50                  55                  60

ATA GAT GGA TGG TAC GGT TAT CAT CAT CAG AAT GAA CAG GGA TCA GGC       240
Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn Glu Gln Gly Ser Gly
65                  70                  75                  80

TAT GCA GCG GAT CAA AAA AGC ACA CAA AAT GCC ATT AAC GGG ATT ACA       288
Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala Ile Asn Gly Ile Thr
                85                  90                  95

AAC AAG GTG AAC TCT GTT ATC GAG AAA ATG AAC ATT CAA TTC ACA GCT       336
Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn Ile Gln Phe Thr Ala
            100                 105                 110

GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG ATG GAA AAT TTA AAT       384
Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu Asn
        115                 120                 125

AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA TAT AAT GCA GAA       432
Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala Glu
    130                 135                 140

TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT TTC CAT GAC TCA       480
Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp Ser
145                 150                 155                 160
```

```
AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA TTA AAG AAT AAT      528
Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn Asn
                165                 170                 175

GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC CAC AAG TGT GAC      576
Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys Asp
            180                 185                 190

AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT GAT TAT CCC AAA      624
Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro Lys
        195                 200                 205

TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA GAT GGA GTG AAA      672
Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val Lys
    210                 215                 220

TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC TAC TCA ACT GTC      720
Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val
225                 230                 235                 240

GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA ATC AGT TTC TGG      768
Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe Trp
                245                 250                 255

ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC ATC TGA              810
Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
            260                 265
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 269 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met Asp Leu Ser Arg Gly
        35                  40                  45

Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly Met
    50                  55                  60

Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn Glu Gln Gly Ser Gly
65                  70                  75                  80

Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala Ile Asn Gly Ile Thr
                85                  90                  95

Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn Ile Gln Phe Thr Ala
            100                 105                 110
```

```
Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu Asn
        115                 120                 125

Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala Glu
    130                 135                 140

Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp Ser
145                 150                 155                 160

Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn Asn
                165                 170                 175

Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys Asp
            180                 185                 190

Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro Lys
        195                 200                 205

Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val Lys
    210                 215                 220

Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val
225                 230                 235                 240

Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe Trp
                245                 250                 255

Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
            260                 265
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 630 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..627

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG        48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC        96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC ATG GAT CAT ATG TTA ACA       144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met Asp His Met Leu Thr
        35                  40                  45

AGT ACT CGA TCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG ATG       192
Ser Thr Arg Ser Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met
    50                  55                  60
```

```
GAA AAT TTA AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA        240
Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr
 65                  70                  75                  80

TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT        288
Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp
                 85                  90                  95

TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA        336
Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln
            100                 105                 110

TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC        384
Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr
        115                 120                 125

CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT        432
His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr
    130                 135                 140

GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA        480
Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val
145                 150                 155                 160

GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC        528
Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile
                165                 170                 175

TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA        576
Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala
            180                 185                 190

ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC        624
Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys
        195                 200                 205

ATC TGA                                                               630
Ile
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 209 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met Asp His Met Leu Thr
        35                  40                  45

Ser Thr Arg Ser Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met
    50                  55                  60
```

```
Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr
 65                  70                  75                  80

Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp
                85                  90                  95

Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln
            100                 105                 110

Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr
        115                 120                 125

His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr
    130                 135                 140

Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val
145                 150                 155                 160

Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile
                165                 170                 175

Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala
            180                 185                 190

Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys
        195                 200                 205

Ile
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..714

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG      48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC      96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC     144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA     192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60
```

```
GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC      240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

ATG CAG ATC CCG GAA TTC AAC AAA TTA GAA AAA AGG ATG GAA AAT TTA      288
Met Gln Ile Pro Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu
                 85                  90                  95

AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA TAT AAT GCA      336
Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala
            100                 105                 110

GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT TTC CAT GAC      384
Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp
        115                 120                 125

TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA TTA AAG AAT      432
Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn
    130                 135                 140

AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC CAC AAG TGT      480
Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys
145                 150                 155                 160

GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT GAT TAT CCC      528
Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
                165                 170                 175

AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA GAT GGA GTG      576
Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
            180                 185                 190

AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC TAC TCA ACT      624
Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
        195                 200                 205

GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA ATC AGT TTC      672
Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
    210                 215                 220

TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC ATC              714
Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
225                 230                 235

TGA                                                                 717
```

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 238 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30
```

```
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu
                85                  90                  95

Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala
            100                 105                 110

Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp
        115                 120                 125

Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn
    130                 135                 140

Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys
145                 150                 155                 160

Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
                165                 170                 175

Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
            180                 185                 190

Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
        195                 200                 205

Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
    210                 215                 220

Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
225                 230                 235
```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 681 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..678

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG          48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC          96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC         144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA         192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC         240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

ATG CAG ATC CCG AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG         288
Met Gln Ile Pro Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
                85                  90                  95

ACA TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG         336
Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
            100                 105                 110

GAT TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC         384
Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
        115                 120                 125

CAA TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC         432
Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
    130                 135                 140

TAC CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT         480
Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
145                 150                 155                 160

TAT GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG         528
Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
                165                 170                 175

GTA GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG         576
Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
            180                 185                 190

ATC TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG         624
Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
        195                 200                 205

GCA ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA         672
Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
    210                 215                 220

TGC ATC TGA                                                            681
Cys Ile
225
```

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 226 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
                85                  90                  95

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
            100                 105                 110

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
        115                 120                 125

Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
    130                 135                 140

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
145                 150                 155                 160

Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
                165                 170                 175

Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
            180                 185                 190

Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
        195                 200                 205

Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
    210                 215                 220

Cys Ile
225
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 158 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Val Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
                85                  90                  95

Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
            100                 105                 110

Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
        115                 120                 125

Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
    130                 135                 140

Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 163 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45
```

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
50 55 60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65 70 75 80

Met Asp Leu Ser Arg Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu
85 90 95

Gly Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln
100 105 110

Asn Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn
115 120 125

Ala Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met
130 135 140

Asn Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Ser Cys Leu Thr Ala
145 150 155 160

Tyr His Arg

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 231 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1 5 10 15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
20 25 30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
35 40 45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
50 55 60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65 70 75 80

Met Gln Ile Pro Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
85 90 95

Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
100 105 110

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
115 120 125

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
130 135 140

```
Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
145                 150                 155                 160

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
                165                 170                 175

Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
            180                 185                 190

Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
        195                 200                 205

Ile Tyr Ser Thr Val Ala Ser Ser Gly Gly Ser Tyr Ser Met Glu His
    210                 215                 220

Phe Arg Trp Gly Lys Pro Val
225                 230
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 225 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1               5                   10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
                85                  90                  95

Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
            100                 105                 110

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
        115                 120                 125

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
    130                 135                 140

Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
145                 150                 155                 160

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
                165                 170                 175
```

```
Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
            180                 185                 190

Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
        195                 200                 205

Ile Tyr Ser Thr Val Ala Ser Ser Gly Gly Ser Tyr Ser Met Leu Val
    210                 215                 220

Asn
225
```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 912 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..912

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG        48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC        96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC       144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA       192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC       240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

ATG CAG ATC CCG GGT CTA TTT GGA GCC ATT GCC GGT TTT ATT GAA GGG       288
Met Gln Ile Pro Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly
                85                  90                  95

GGA TGG ACT GGA ATG ATA GAT GGA TGG TAC GGT TAT CAT CAT CAG AAT       336
Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn
            100                 105                 110

GAA CAG GGA TCA GGC TAT GCA GCG GAT CAA AAA AGC ACA CAA AAT GCC       384
Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala
        115                 120                 125
```

```
ATT AAC GGG ATT ACA AAC AAG GTG AAC TCT GTT ATC GAG AAA ATG AAC        432
Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn
    130                 135                 140

ATT CAA TTC ACA GCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG        480
Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
145                 150                 155                 160

ATG GAA AAT TTA AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG        528
Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
                165                 170                 175

ACA TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG        576
Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
            180                 185                 190

GAT TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC        624
Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
        195                 200                 205

CAA TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC        672
Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
    210                 215                 220

TAC CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT        720
Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
225                 230                 235                 240

TAT GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG        768
Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
                245                 250                 255

GTA GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG        816
Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
            260                 265                 270

ATC TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG        864
Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
        275                 280                 285

GCA ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA        912
Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
    290                 295                 300
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 304 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30
```

```
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly
                85                  90                  95

Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn
            100                 105                 110

Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala
        115                 120                 125

Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn
    130                 135                 140

Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
145                 150                 155                 160

Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
                165                 170                 175

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
            180                 185                 190

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
        195                 200                 205

Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
    210                 215                 220

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
225                 230                 235                 240

Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
                245                 250                 255

Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
            260                 265                 270

Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
        275                 280                 285

Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
    290                 295                 300
```

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..471

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG ATG GAA AAT TTA AAT        48
Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu Asn
 1               5                  10                  15

AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA TAT AAT GCA GAA        96
Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala Glu
            20                  25                  30

TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT TTC CAT GAC TCA       144
Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp Ser
        35                  40                  45

AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA TTA AAG AAT AAT       192
Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn Asn
    50                  55                  60

GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC CAC AAG TGT GAC       240
Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys Asp
65                  70                  75                  80

AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT GAT TAT CCC AAA       288
Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro Lys
                85                  90                  95

TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA GAT GGA GTG AAA       336
Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val Lys
            100                 105                 110

TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC TAC TCA ACT GTC       384
Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val
        115                 120                 125

GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA ATC AGT TTC TGG       432
Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe Trp
    130                 135                 140

ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC ATC TGA               474
Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 157 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu Asn
 1               5                  10                  15
```

```
Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala Glu
            20                  25                  30

Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp Ser
        35                  40                  45

Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn Asn
    50                  55                  60

Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys Asp
65                  70                  75                  80

Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro Lys
                85                  90                  95

Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val Lys
            100                 105                 110

Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val
        115                 120                 125

Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe Trp
    130                 135                 140

Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CATGGATCAT ATGTTAACAG ATATCAAGGC CTGACTGACT GAGAGCT 47

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CTAGTATACA ATTGTCTATA GTTCCGGACT GACTGACTC 39

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CATGGGCGCC CATATGGGCA TATTCGGCG 29

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCGCGGGTAT ACCCGTATAA GCC 23

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CATGGATCAT ATGTTAACAA GTACTCGATA TCAATGAGTG ACTGAAGCT 49

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CTAGTATACA ATTGTTCATG AGCTATAGTT ACTCACTGAC T 41

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AATTCGTACC TA 12

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCATGGATCT AG 12

WHAT IS CLAIMED IS:

1. A vaccine for stimulating protection in animals against infection by influenza virus which comprises a an effective amount of an immunogenic fragment of the HA2 subunit of an HA protein selected from the group consisting of a type A subtype influenza virus or a type B influenza virus.

2. The vaccine according to claim 1 wherein said type A subunit is H3N2.

3. The vaccine according to claim 1 wherein the polypeptide is fused to a second polypeptide.

4. The vaccine according to claim 2 wherein the second polypeptide comprises the N terminal amino acids of a NS1 protein.

5. The vaccine according to claim 1 wherein the immunogenic fragment of the HA2 subunit is selected from the group consisting of a peptide comprising amino acids 1 to 221 of the H3HA2 subtype, a peptide comprising amino acids 77 to 221 of the H3HA2 subtype, a peptide comprising amino acids 1 to 223 of the BHA2 type, and a peptide comprising amino acids 41 to 223 of the BHA2 type.

6. The vaccine according to claim 5 comprising $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10.

7. The vaccine according to claim 5 comprising $NS1_{(1-81)}H3HA2_{(77-221)}$ SEQ ID NO: 12.

8. The vaccine according to claim 5 comprising $NS1_{1-42}BLHA2_{41-223}$ SEQ ID NO: 14.

9. A protein comprising an immunogenic fragment of the HA2 subunit of an HA protein selected from the group consisting of Type A subtype or type B influenza virus.

10. The protein according to claim 9 wherein said type A subtype is H3N2.

11. The protein according to claim 9 wherein the peptide containing the immunogenic fragment is fused to a second peptide or protein.

12. The protein according to claim 10 wherein the second peptide comprises the N terminal amino acids of a NS1 protein.

13. The protein according to claim 10 wherein the immunogenic fragment of the HA2 subunit is selected from the group consisting of a peptide comprising amino acids 1 to 221 of the H3HA2 subunit, a peptide comprising amino acids 77 to 221 of the H3HA2 subunit, a peptide comprising amino acids 1-223 of the BHA2 subunit, and a peptide comprising amino acids 41-223 of the BHA2 subunit.

14. A polypeptide $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10.

15. A polypeptide $NS1_{(1-81)}H3HA2_{(77-221)}$ SEQ ID NO: 12.

16. A polypeptide $NS1_{1-41}BLHA2_{41-223}$ SEQ ID NO: 14.

17. A DNA molecule comprising a coding sequence for an immunogenic fragment of the HA2 subunit of an HA protein selected from the group consisting of a Type A subtype or type B influenza virus.

18. The DNA molecule according to claim 17 wherein said Type A subunit is H3N2.

19. The DNA molecule according to claim 17 comprising a coding sequence for the polypeptide $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10.

20. The DNA molecule according to claim 17 comprising a coding sequence for the polypeptide $NS1_{(1-42)}H3BLHA2_{(41-223)}$ SEQ ID NO: 14.

21. The DNA molecule according to claim 17 comprising a coding sequence for the polypeptide $NS1_{(1-81)}H3HA2_{(77-221)}$ SEQ ID NO: 12.

22. Plasmid pMG13H3HA SEQ ID NO: 9.

23. Plasmid $pNS1_{1-41}BLHA2_{41-223}$ SEQ ID NO: 13.

24. A microorganism transformed with a DNA molecule comprising a coding sequence for an immunogenic fragment of the HA2 subunit of an HA protein selected from the group consisting of a Type A subtype or type B influenza virus.

25. The microorganism according to claim 24 wherein said Type A subunit is H3N2.

26. The microorganism according to claim 24 wherein said DNA molecule comprises a coding sequence for the polypeptide $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10.

27. A combination vaccine for stimulating protection in animals against infection by influenza virus which comprises a first polypeptide having an immunogenic fragment of the HA2 subunit of an influenza H3 subtype virus and a second polypeptide selected from the group consisting of a polypeptide having an immunogenic fragment of the HA2 subunit of a type B influenza virus, and a polypeptide having an immunogenic fragment of the HA2 subunit of an H1 subtype influenza virus, and a polypeptide having an immunogenic fragment of the HA2 subunit of an H2 subtype influenza virus.

28. The combination vaccine according to claim 27 wherein the first polypeptide is selected from the group consisting of $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10 and $NS1_{(1-81)}H3HA2_{(77-221)}$ SEQ ID NO: 12.

29. The combination vaccine according to claim 27 wherein the second polypeptide is a polypeptide having an immunogenic fragment of the HA2 subunit of an H1 subtype influenza virus.

30. The combination vaccine according to claim 27 wherein said second polypeptide is selected from the group consisting of C13 SEQ ID NO: 16, D SEQ ID NO: 18, C13 short SEQ ID NO: 20, D short SEQ ID NO: 22, A SEQ ID NO: 24, C SEQ ID NO: 26, ΔD SEQ ID NO: 27, Δ13 SEQ ID NO: 28, M SEQ ID NO: 29, ΔM SEQ ID NO: 30, ΔM+ SEQ ID NO: 32, and $H1HA2_{66-222}$ SEQ ID NO: 34.

31. The combination vaccine according to claim 27 wherein said second polypeptide is $NS1_{1-42}BLHA2_{41-223}$ SEQ ID NO: 14.

32. A combination vaccine for stimulating protection in animals against infection by influenza virus which comprises a first polypeptide having an immunogenic fragment of the HA2 subunit of an influenza H3 subtype virus, a second polypeptide having an immunogenic fragment of the HA2 subunit of an influenza B type virus, and a third polypeptide selected from the group consisting of a polypeptide having an immunogenic fragment of the HA2 subunit of an H1 subtype influenza virus and a polypeptide having an immunogenic fragment of the HA2 subunit of an H2 subtype influenza virus.

FIGURE 1

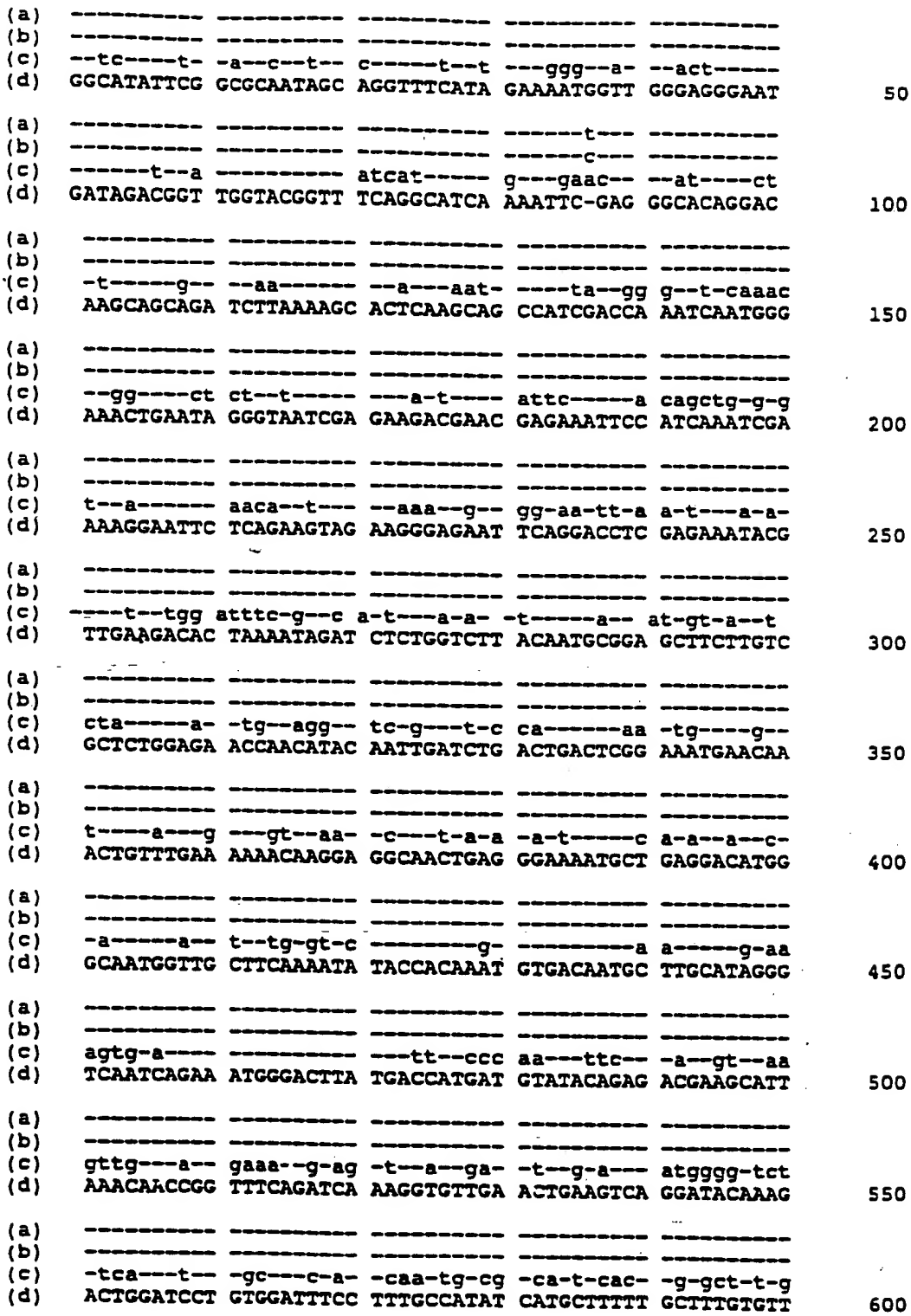
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) --tc----t- -a--c--t-- c-----t--t ---ggg--a- --act-----
(d) GGCATATTCG GCGCAATAGC AGGTTTCATA GAAAATGGTT GGGAGGGAAT 50
(a) ---------- ---------- ---------- ------t--- ----------
(b) ---------- ---------- ---------- ------c--- ----------
(c) ------t--a ---------- atcat----- g---gaac-- --at----ct
(d) GATAGACGGT TGGTACGGTT TCAGGCATCA AAATTC-GAG GGCACAGGAC 100
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) -t-----g-- --aa------ --a---aat- ----ta--gg g--t-caaac
(d) AAGCAGCAGA TCTTAAAAGC ACTCAAGCAG CCATCGACCA AATCAATGGG 150
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) --gg----ct ct--t----- ---a-t---- attc-----a cagctg-g-g
(d) AAACTGAATA GGGTAATCGA GAAGACGAAC GAGAAATTCC ATCAAATCGA 200
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) t--a------ aaca--t--- --aaa--g-- gg-aa-tt-a a-t---a-a-
(d) AAAGGAATTC TCAGAAGTAG AAGGGAGAAT TCAGGACCTC GAGAAATACG 250
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) ----t--tgg atttc-g--c a-t---a-a- -t-----a-- at-gt-a--t
(d) TTGAAGACAC TAAAATAGAT CTCTGGTCTT ACAATGCGGA GCTTCTTGTC 300
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) cta-----a- -tg--agg-- tc-g---t-c ca------aa -tg----g--
(d) GCTCTGGAGA ACCAACATAC AATTGATCTG ACTGACTCGG AAATGAACAA 350
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) t----a---g ---gt--aa- -c---t-a-a -a-t-----c a-a--a--c-
(d) ACTGTTTGAA AAAACAAGGA GGCAACTGAG GGAAAATGCT GAGGACATGG 400
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) -a-----a-- t--tg-gt-c --------g- ---------a a-----g-aa
(d) GCAATGGTTG CTTCAAAATA TACCACAAAT GTGACAATGC TTGCATAGGG 450
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) agtg-a---- ---------- ---tt--ccc aa---ttc-- -a--gt--aa
(d) TCAATCAGAA ATGGGACTTA TGACCATGAT GTATACAGAG ACGAAGCATT 500
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) gttg---a-- gaaa--g-ag -t--a--ga- -t--g-a--- atgggg-tct
(d) AAACAACCGG TTTCAGATCA AAGGTGTTGA ACTGAAGTCA GGATACAAAG 550
(a) ---------- ---------- ---------- ---------- ----------
(b) ---------- ---------- ---------- ---------- ----------
(c) -tca---t-- -gc---c-a- -caa-tg-cg -ca-t-cac- -g-gct-t-g
(d) ACTGGATCCT GTGGATTTCC TTTGCCATAT CATGCTTTTT GCTTTGTGTT 600

FIGURE 1 (con't)

```
(a)  ---------- ---------. ..-------- -----g---- ----------
(b)  ---------- ---------. ..-------- -----a---- ----------
(c)  --c-cc---- --gca----g tt-c---atg --ttct--t- -atctt-gca
(d)  GTTTTGCTGG GGTTCATCA- --TGTGGGCC TGCCA-AAAG GCAACATTAG      650

(a)  ---------- ---------.
(b)  ---------- ---------.
(c)  -----ga--a -----c---g
(d)  GTGCAACATT TGCATTTGA-                                       670
```

FIGURE 2

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG          48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC          96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC         144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA         192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC         240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

ATG GGC GCC CAT ATG GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA         288
Met Gly Ala His Met Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu
                85                  90                  95

AAT GGT TGG GAG GGA ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT CAA         336
Asn Gly Trp Glu Gly Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln
            100                 105                 110

AAT TCT GAG GGC ACA GGA CAA GCA GCA GAT CTT AAA AGC ACT CAA GCA         384
Asn Ser Glu Gly Thr Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala
        115                 120                 125

GCC ATC GAC CAA ATC AAT GGG AAA CTG AAT AGG GTA ATC GAG AAG ACG         432
Ala Ile Asp Gln Ile Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr
    130                 135                 140

AAC GAG AAA TTC CAT CAA ATC GAA AAG GAA TTC TCA GAA GTA GAA GGG         480
Asn Glu Lys Phe His Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly
145                 150                 155                 160

AGA ATT CAG GAC CTC GAG AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC         528
Arg Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu
                165                 170                 175

TGG TCT TAC AAT GCG GAG CTT CTT GTC GCT CTG GAG AAC CAA CAT ACA         576
Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr
            180                 185                 190

ATT GAT CTG ACT GAC TCG GAA ATG AAC AAA CTG TTT GAA AAA ACA AGG         624
Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg
        195                 200                 205

AGG CAA CTG AGG GAA AAT GCT GAG GAC ATG GGC AAT GGT TGC TTC AAA         672
Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys
    210                 215                 220
```

FIGURE 2 (con't)

```
ATA TAC CAC AAA TGT GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG          720
Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly
225                 230                 235                 240

ACT TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA AAC AAC CGG TTT          768
Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe
                    245                 250                 255

CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC AAA GAC TGG ATC CTG          816
Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu
            260                 265                 270

TGG ATT TCC TTT GCC ATA TCA TGC TTT TTG CTT TGT GTT GTT TTG CTG          864
Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu
        275                 280                 285

GGG TTC ATC ATG TGG GCC TGC CAA AAA GGC AAC ATT AGG TGC AAC ATT          912
Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile
    290                 295                 300

TGC ATT                                                                 918
Cys Ile
305
```

FIGURE 3

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG          48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC          96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC         144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA         192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC         240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

ATG GAT CAT ATG TTA ATT CAG GAC CTC GAG AAA TAC GTT GAA GAC ACT         288
Met Asp His Met Leu Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr
                85                  90                  95

AAA ATA GAT CTC TGG TCT TAC AAT GCG GAG CTT CTT GTC GCT CTG GAG         336
Lys Ile Asp Leu Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu
            100                 105                 110

AAC CAA CAT ACA ATT GAT CTG ACT GAC TCG GAA ATG AAC AAA CTG TTT         384
Asn Gln His Thr Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe
        115                 120                 125

GAA AAA ACA AGG AGG CAA CTG AGG GAA AAT GCT GAG GAC ATG GGC AAT         432
Glu Lys Thr Arg Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn
    130                 135                 140

GGT TGC TTC AAA ATA TAC CAC AAA TGT GAC AAT GCT TGC ATA GGG TCA         480
Gly Cys Phe Lys Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser
145                 150                 155                 160

ATC AGA AAT GGG ACT TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA         528
Ile Arg Asn Gly Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu
                165                 170                 175

AAC AAC CGG TTT CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC AAA         576
Asn Asn Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys
            180                 185                 190

GAC TGG ATC CTG TGG ATT TCC TTT GCC ATA TCA TGC TTT TTG CTT TGT         624
Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys
        195                 200                 205

GTT GTT TTG CTG GGG TTC ATC ATG TGG GCC TGC CAA AAA GGC AAC ATT         672
Val Val Leu Leu Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile
    210                 215                 220

AGG TGC AAC ATT TGC ATT                                                690
Arg Cys Asn Ile Cys Ile
225                 230
```

FIGURE 4

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TCC TTT CTT TGG          48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Ser Phe Leu Trp
 1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC          96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC ATG CAT GGA TCA TAT GTT         144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met His Gly Ser Tyr Val
        35                  40                  45

AAC AAG ACA CAA GAA GCT ATA AAC AAG ATA ACA AAA AAT CTC AAC TAT         192
Asn Lys Thr Gln Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Tyr
    50                  55                  60

TTA AGT GAG CTA GAA GTA AAA AAC CTT CAA AGA CTA AGC GGA GCA ATG         240
Leu Ser Glu Leu Glu Val Lys Asn Leu Gln Arg Leu Ser Gly Ala Met
65                  70                  75                  80

AAT GAG CTT CAC GAC GAA ATA CTC GAG CTA GAC GAA AAA GTG GAT GAT         288
Asn Glu Leu His Asp Glu Ile Leu Glu Leu Asp Glu Lys Val Asp Asp
                85                  90                  95

CTA AGA GCT GAT ACA ATA AGC TCA CAA ATA GAG CTT GCA GTC TTG CTT         336
Leu Arg Ala Asp Thr Ile Ser Ser Gln Ile Glu Leu Ala Val Leu Leu
            100                 105                 110

TCC AAC GAA GGG ATA ATA AAC AGT GAA GAT GAG CAT CTC TTG GCA CTT         384
Ser Asn Glu Gly Ile Ile Asn Ser Glu Asp Glu His Leu Leu Ala Leu
        115                 120                 125

GAA AGA AAA CTG AAG AAA ATG CTT GGC CCC TCT GCT GTA GAA ATA GGG         432
Glu Arg Lys Leu Lys Lys Met Leu Gly Pro Ser Ala Val Glu Ile Gly
    130                 135                 140

AAT GGG TGC TTT GAA ACC AAA CAC AAA TGC AAC CAG ACT TGC CTA GAC         480
Asn Gly Cys Phe Glu Thr Lys His Lys Cys Asn Gln Thr Cys Leu Asp
145                 150                 155                 160

AGG ATA GCT GCT GGC ACC TTT AAT GCA GGA GAT TTT TCT CTT CCC ACT         528
Arg Ile Ala Ala Gly Thr Phe Asn Ala Gly Asp Phe Ser Leu Pro Thr
                165                 170                 175

TTT GAT TCA TTA AAC ATT ACT GCT GCA TCT TTA AAT GAT GAT GGC TTG         576
Phe Asp Ser Leu Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu
            180                 185                 190
```

FIGURE 4 (con't)

```
GAT AAT CAT ACT ATA CTG CTC TAC TAC TCA ACT GCT GCT TCT AGC TTG          624
Asp Asn His Thr Ile Leu Leu Tyr Tyr Ser Thr Ala Ala Ser Ser Leu
        195                 200                 205

GCT GTA ACA TTA ATG ATA GCT ATC TTC ATT GTC TAC ATG GTC TCC AGA          672
Ala Val Thr Leu Met Ile Ala Ile Phe Ile Val Tyr Met Val Ser Arg
    210                 215                 220

GAC AAT GTT TCT TGT TCC ATC TGT CTG                                      699
Asp Asn Val Ser Cys Ser Ile Cys Leu
225                 230
```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/01451

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :IPC:5 A61K 39/12;CO7K 3/00; CO7H 15/12
US CL :424/89; 530/350; 536/27
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/89; 530/350; 536/27

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Dialog, APS, search terms: influenza virus, hemagglutinin, nonstructural protein, fusion protein, vaccine

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Journal of Experimental Medicine, Volume 162, issued August, 1985, Yamada et al, "Influenza Virus Hemagglutinin-Specific Cytotoxic T Cell Response Induced By Polypeptide Produced In Escherichia coli", pages 663-674, especially pages 664-665.	1-32
X	Journal Of Experimental Medicine, volume 162, issued November 1985, Yamada et al, "Influenza virus Subtype-specific Cytotoxic T Lymphocytes Lyse Target CEIls Coated With A Protein Produced In E. coli", pages 1720-1725, see entire document.	1-16, 27-32

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
06 May 1993

Date of mailing of the international search report
17 MAY 1993

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. NOT APPLICABLE

Authorized officer
L. F. SMITH
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/01451

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Journal Of Immunology, volume 140, No. 4, issued 15 February 1988, Kuwano et al, "HA2 Subunit Of Influenza A H1 and H2 Subtype Viruses Induces A Protective Cross-Reactive Cytotoxic T Lymphocyte Response", pages 1264-1268, see entire document.	1-16, 27-32
Y	Federation of American Societies For Experimental Biology, 75th Anual Meeting, volume 5, no. 5, issued 21-25 April, 1991, Dillon, et al, "Activity of CD8+ CTL In Mice Immunized With Recombinant Influenza NS1-HA2 Fusion Protein Or A CTL Epitope Peptide (HA2 189-199), Abstract 5748, page A1362.	1-16, 27-32